主编

郝柏林

刘寄星

生命科学丛书



上海科学技术出版社

理论物理与生命科学



郝柏林

1934年6月26日生，中国科学院院士。1959年6月毕业于苏联哈尔科夫大学，1960—1963年在莫斯科大学作研究生。先后任中国科学院物理研究所研究员，理论物理研究所所长(1991—1994)。曾在固体物理、量子多体问题、相变和重整化群、天线理论、非平衡统计物理、非线性系统的混沌现象、计算物理等多个领域开展了大量的研究工作，曾多次应邀在重要国际会议上作邀请报告及在国外科研机构访问、工作。在国内外出版有关计算物理、相变理论和混沌理论等领域的著作十几种。他的研究成果得到国内外同行的承认和重视，曾获1978年中国科学院重大成果奖，1987年中国科学院科学进步二等奖，1992年中国科学院自然科学一等奖，1993年国家自然科学二等奖。

刘寄星

1938年11月生，研究员。1964年毕业于北京大学物理系，先后在二机部401所、七机部207所从事国防科研工作。1985年获美国得克萨斯大学物理学博士学位，1986年起在中国科学院理论物理研究所工作，现任理论物理研究所副所长。从事高温气体输运性质、等离子体与飞行体相互作用、电波传播、磁化等离子体的稳定性和非线性等离子体理论方面的研究。主要研究方向为非线性等离子体理论和转动流体相干结构，1990年参加生物膜形状的液晶模型理论研究，获1995年中国科学院自然科学一等奖。



封面设计　陈　蕾

ISBN 7-5323-4398-7

9 787532 343980 >

定　价：30.00元

生命科学丛书

理论物理与生命科学

郝柏林　刘寄星　主编

上 海 科 学 技 术 出 版 社

丛书策划：濮紫兰

生命科学丛书
理论物理与生命科学
郝柏林　刘寄星　主编
上海科学技术出版社出版、发行
（上海瑞金二路450号　邮政编码200020）
新华书店 上海发行所经销
商务印书馆 上海印刷股份有限公司印刷
开本787×1092　小1/16　印张19　插页4　字数281 000
1997年12月第1版　1999年11月第2次印刷
印数1 201—2 200
ISBN 7-5323-4398-7/Q·65
定价：30.00元

《生命科学丛书》序

本世纪50年代DNA双螺旋结构的发现，以及随后遗传信息传递的中心法则的确定、重组DNA技术的建立，分子生物学和遗传学的概念和技术已渗透到生命科学的各个分支学科中，生命科学与物理学、化学、数学的交叉渗透，计算机和大量新技术的广泛应用，已使当代生命科学的面貌发生了根本性的变化。科学家预言，21世纪将是生命科学的世纪。

生命科学的研究对象是形形色色的生物体及其复杂的生命活动的规律。与生物的多样性一样，生命科学的分支学科也是多种多样的。有根据研究的生物对象区分的，如植物学、动物学、微生物学、病毒学等；也有按研究的生命活动过程或层次区分的，如分子生物学、生物化学、细胞生物学、遗传学、发育生物学、生理学、形态学、生态学等。当代生命科学的特点是：由于学科交叉的结果，生命科学的发展极为迅速，传统的学科界限正在逐渐消失，而新的分支学科或生长点又在不断产生，如由于人类基因组和其他生物基因组研究而产生的基因组学，由利用分子生物学手段研究生物系统和进化而产生的分子系统学等。同时，生命科学又与人类的日常生活密切相关，它直接关系到人类自身及其生存环境的维护，如人口控制、疾病的预防和治疗、农产品的生产、生物多样性和环境的保护等。由于生命科学是当代医学和农学的重要基础，在很多研究领域，基础研究和应用研究之间的距离正在缩短。特别是分子生物学的发展，以重组DNA技术为代表的生物工程的兴起，已为近代生物医药产业的发展、重大疾病的预防和治疗、农作物的改良及病虫害防治的研究和实践注入了新的动力，提供了一系列新的技术。在这些领域中不断传来令人振奋的消息。

至今，有科学记载的生物已有170多万种，最保守的估计全球也有500万～1 000万种生物。地球上有不少地方特别是热带地区，还没有作过系统的调查研究，因此地球上物种的数目也可能远远超过此数。系统进化的研究，生物多样性的研究，将为人们认识生物演化的历史进程，生物多

样性的保护，以及自然资源的合理开发和利用，提供扎实的科学基础。

为了反映当代生命科学的研究进展及其在社会经济发展中的作用和影响，并展现我国科学家所取得的研究成果，上海科学技术出版社就生命科学中的若干专题，聘请有关专家学者编著这套《生命科学丛书》，力求从一个个侧面，向不同层次的读者介绍有关的生物学知识，及其实际应用的现况或潜力，并以此促进社会各界更好地关心和支持我国生命科学的发展，吸引更多的青年人投身到生命科学及其相关的研究行列中来，在中华大地上为发展我国的生命科学及生物工程作出贡献。让我们面对新世纪的曙光，迎接生命科学世纪的到来。

许智宏

（《生命科学丛书》主编）

1997 年 7 月

谨以本书献给

彭桓武教授八十华诞！

撰稿人名单

（以所撰文稿出现先后排序）

邹承鲁	中国科学院生物物理研究所，中科院院士（北京 100101）
郝柏林	中国科学院理论物理研究所，中科院院士（北京 100080）
徐京华	中国科学院上海生物化学研究所（上海 200031）
阎隆飞	北京农业大学，中科院院士（北京 100094）
彭征宇	Whitehead Institute for Biomedical Research, MIT, Cambridge, MA02142, USA
郑　强	Division of Structural Biology, Scios Nova Inc., Sunnyvale, CA94086, USA
陈润生	中国科学院生物物理研究所（北京 100101）
施蕴渝	中国科学技术大学生物系（合肥 230026）
陈凯先	中国科学院上海药物研究所（上海 200031）
蒋华良	中国科学院上海药物研究所（上海 200031）
唐　赟	中国科学院上海药物研究所（上海 200031）
丁达夫	中国科学院上海生物化学研究所（上海 200031）
汤海旭	中国科学院上海生物化学研究所（上海 200031）
张保红	中国科学院上海生物化学研究所（上海 200031）
陈　洁	中国科学院上海生物化学研究所（上海 200031）
孙之荣	清华大学生物系（北京 100084）
张春霆	天津大学物理系，中科院院士（天津 300072）
罗辽复	内蒙古大学物理系（呼和浩特 010021）
彭守礼	云南大学物理系（昆明 650091）
刘次全	云南大学现代生物研究中心（昆明 650091）
谢惠民	苏州大学数学系（苏州 215006）
林克椿	北京医科大学生物物理系（北京 100083）
欧阳钟灿	中国科学院理论物理研究所（北京 100080）
刘寄星	中国科学院理论物理研究所（北京 100080）
郑伟谋	中国科学院理论物理研究所（北京 100080）

漆安慎　北京师范大学物理系(北京 100875)
曲　直　中国医科大学生物物理教研室(沈阳 110001)
张广才　中国医科大学生物物理教研室(沈阳 110001)
卓益忠　原子能科学研究院核物理研究所(北京 102413)
包景东　北京气象学院(北京 100081)
吴锡真　原子能科学研究院核物理研究所(北京 102413)
彭桓武　中国科学院理论物理研究所,中科院院士(北京 100080)
黄祖洽　北京师范大学低能核物理研究所,中科院院士(北京 100875)

前　　言

物理科学和生命科学的相互作用由来已久。从两门科学发展的历史来看，电学与电生理学几乎是同时诞生的(C. A. de Coulomb，1785；L. Galvani，1791)。生物学为物理学启示了能量守恒定律，物理学则向生物学提供了从显微镜、X 光、示踪原子、中子衍射、核磁共振、同步辐射，一直到扫描隧道显微镜等各种研究手段。然而，只有到了把生命过程和生命物质作为真正的物理对象加以研究之时，理论物理的思维方法和数学工具才会在生物学研究中发挥更大作用。现在，生命科学积累的大量事实和数据，以及物理科学在研究复杂系统方面取得的长足进展，已经把物理学和生命科学密切结合的任务提上了日程。有理由相信，两门科学在新基础之上的进一步结合，必将为这两门科学本身乃至整个自然科学的发展，带来前所未有的崭新局面。

为了积极推动这一结合的进行，即一方面促进我国部分理论物理工作者向生命科学领域作战略进军，另一方面促进生物学工作者与物理学工作者增加交流，1995 年 9 月 28 日至 30 日在北京香山举行了以理论物理与生命科学的结合为主题的学术讨论会(第 40 次香山会议)。来自国内及海外的 18 个科研机构和高等院校的 32 位学者，其中包括 6 位中国科学院院士，就理论物理与生命科学结合中的问题开展了讨论，并就蛋白质与生物大分子、基因序列、生物膜和生物数学模型 4 个具体专题进行了交流。

本书是从与会学者在会议上的专题报告和若干书面发言中加以选择，并约请作者撰稿，汇集而成的。全书共分 5 章，分别从生物学家和理论物理学家的角度阐述理论物理与生命科学相结合的重要性和可能性，讨论了蛋白质和生物大分子、基因序列、生物膜蛋白和生物膜模型、肿瘤生长、心脑可激媒质和分子马达等重要问题。应当指出，本书所包含的内容，除少量综述和介绍性内容外，多数章节都是撰稿人近年来在理论物理学

与生命科学相结合的研究中，所取得的创造性成果。我们希望本书的出版，对理论物理学工作者和生物学工作者加强相互交流有所裨益，也能为有志投身生命科学研究的物理学工作者提供一本反映研究前沿动态的参考资料。

第 40 次香山会议召开时，适逢多年来热情倡导物理科学与生命科学结合的彭桓武院士 80 华诞，会议特邀黄祖洽院士就彭先生对我国科学和教育事业的贡献作了专题发言。为了表达对彭先生促进物理学与生命科学结合所作特殊贡献的感谢并介绍他的学术观点，征得《科学》杂志编辑部同意，我们将彭先生已发表于《科学》杂志的"八十自述"一文，连同黄祖洽先生的讲话一起，作为附录收入本书，以飨读者。

最后，我们要感谢上海科学技术出版社把本书收入该社的生命科学丛书出版。借本书出版之机，我们感谢香山会议委员会的支持，感谢中国科学院物理研究所胡占宁博士在本书组稿过程中所付出的大量劳动。

郝柏林 刘寄星

1997 年 7 月 31 日于北京西三旗

目　　录

第1章　概　　论

§1.1　生物学走向21世纪

生物学在20世纪取得了巨大的进展。数理科学广泛而又深刻地渗入到生物学中，全面地改变了生物学的面貌，开辟了在分子水平研究生命现象的新学科——分子生物学。分子生物学的研究涉及生命现象最本质的内容，把在各个层次的生命活动有机地联系起来，在新的高度上揭示生命的奥秘。分子生物学渗入生物学每一个分支的结果，全面地推动了生物学的发展。细胞生物学和神经生物学已相继进入分子水平，成为生物学领域内新的主要生长点。但它们已经不是原来的经典学科，而是以分子水平研究为基础的、面貌全新的现代细胞生物学和现代神经生物学。即使最古老的生物分类和进化学，也因为分享了分子生物学的成果而焕发青春。

生物学研究进入分子水平，才得以从本质上去探讨生命活动的规律，因此分子生物学成为当代生命科学基础研究中的前沿，开辟了现代生物学的全新局面。另一方面，它又使生物学以空前主动的势态，对人类物质生产和社会生活产生重大影响。1973年重组DNA获得成功，开创了基因工程。以此为基础，生物技术作为前途远大的高技术产业在世界范围兴起，生物工程将成为现代化的大工业。与此同时，它还极大地推动了医学和农业科学的实践，在这些领域中正展示出广阔的应用前景。它必将又一次引起人类社会和经济生活的革命。

分子生物学的兴起改变了生物学的面貌。

1.1.1　组成生物体的基本物质

分子生物学的兴起首先建立在对组成生物体的基本物质的认识上。构成生物体的物质种类很多，但其中最重要的是蛋白质和核酸。

蛋白质是生命活动的主要承担者，一切生命活动无不与蛋白质有关。新陈代谢是生命活动的主要特征，而构成新陈代谢的所有化学变化都是

在酶的催化下进行的。除最近发现的极少数具有催化功能的核糖核酸外，其他的酶都是蛋白质。生长，运动，呼吸，免疫，消化，光合作用，以及对外界环境变化的感觉并作出必要的反应等，都必须依靠蛋白质来实现。虽然遗传信息的携带者是核酸，但遗传信息的传递和表达不仅仍然是在酶的催化之下，而且也是在蛋白质的调节控制下进行的。

几乎所有的蛋白质都是由 20 种氨基酸以肽键连接而成的。蛋白质中氨基酸的排列顺序称为蛋白质的一级结构。如果你怀疑丰富多采的生命世界怎么可能由仅仅 20 种氨基酸构成的蛋白质所体现，那么一个简单的计算可以回答这个问题。一个仅含有 100 个氨基酸的蛋白质几乎是最小的蛋白质了。但是，在这样的一个小蛋白质中 20 种氨基酸的排列有 20^{100} 或 10^{130} 种不同的方式。也就是说，可以构成 10^{130} 种不同的蛋白质，这是一个巨大的天文数字。即使每一种蛋白质只有一个分子，它的总重量也将达到 10^{100} 吨，这个重量是地球重量的 10^{78} 倍，是太阳系总重量的 10^{72} 倍。不但在地球上生命进化的 40 亿年过程中所有的蛋白质不会超过这个重量，再过 40 亿年也不会超过这个重量。何况这还只不过是考虑了最小的蛋白质而已！

每一个蛋白质都有它自己特定的氨基酸序列和特定的空间结构。50 年代中期，胰岛素分子的氨基酸序列及二硫键连接方式的阐明是蛋白质一级结构测定的开始。40 年来，氨基酸序列被测定的蛋白质已有数千个。蛋白质分子除有以氨基酸组成的并有一定顺序的肽链结构外，还具有肽链在空间的卷曲折叠而形成的三维空间结构。第一个被测定空间结构的蛋白质是肌红蛋白。只有处在这种特定的三维结构中的蛋白质分子才是能够发挥生物功能的活性蛋白质。因此，即使肽链的氨基酸序列不变，只要空间结构被破坏，也会导致蛋白质功能的丧失。蛋白质在肽链保持完整的情况下空间结构的破坏称为蛋白质的变性，这一概念是我国科学家吴宪在 30 年代初根据他在国内的工作首先提出来的，长期以来已被国际上广泛接受。

蛋白质晶体 X 射线衍射仍然是蛋白质空间结构测定的主要方法。已经确定空间结构并存入数据库的蛋白质已达 1 000 个以上，但是这些一级结构和空间结构被测定的蛋白质，只不过是自然界数以百万计的蛋白质中的一个微不足道的部分。在结构研究领域内，近 20 年来发展起来的

二维核磁共振方法已经显示了它对蛋白质在溶液中的空间结构和运动状态方面研究的优势，现已测出了上百个较小蛋白质的结构，也许会在不远的将来为生物大分子空间结构测定带来又一次突破。

结构与功能关系的研究一直是蛋白质研究的核心问题之一。过去最常用的方法是用化学方法修饰蛋白质的侧链基团，以观察对蛋白质生物活性的影响。现在，体外基因突变技术，特别是定点突变的发明，可以任意改变蛋白质分子中的氨基酸残基，并观察其对生物功能的影响。另一个重要的问题是蛋白质空间结构与其生物活性的关系。我们新近的研究结果指出，空间结构对酶的功能至关重要，即使极其细微的扰乱，也会导致酶活力的丧失。但是，蛋白质分子并不是一个刚性分子，它的空间结构在一定程度上是在不断运动之中，即使在晶体状态下运动也不停止。实际上，蛋白质的功能不仅与分子结构本身密切相关，而且必须依赖于结构的这种运动性能。酶分子活性部位的一定程度的柔性，亦即可运动性，正是酶充分发挥其催化功能所必需的。

核酸是遗传信息的承担者，其分子是由许许多多核苷酸相连接的长链。每一个核苷酸包含碱基、核糖和磷酸三部分。根据核糖结构的不同，核酸分为两类，含核糖的和含脱氧核糖的核酸，它们分别称为核糖核酸(RNA)和脱氧核糖核酸(DNA)。虽然碱基种类较多，但DNA和RNA都分别各含四种碱基，构成四种核苷酸。在DNA分子中的碱基是腺嘌呤、鸟嘌呤、胸腺嘧啶和胞嘧啶。RNA分子中的碱基除以尿嘧啶代替胸腺嘧啶外，其他的与DNA中的完全相同。DNA以双股核苷酸链形式存在，在双链之间存在着根据其碱基性质严格的两两配对关系：脱氧腺苷酸和脱氧胸苷酸配对，脱氧鸟苷酸和脱氧胞苷酸配对。复制时双链解离，各以单链为模板，按照上述配对原则，合成一条新的单链，这样就生成了和原来完全相同的两个双股DNA分子。绝大多数生物体的遗传信息存在于DNA分子之中，DNA的复制构成了遗传的分子基础。这一复制过程是极为可靠的，发生错误的可能性虽然仅为一万亿分之一，但还是可能会发生的。如果DNA的复制是绝对可靠，绝对没有任何错误，那么生物体的突变就只能靠外界因素才能引起，生物的进化和生物界的面貌也就不是今天的样子了。

RNA分子也可以与DNA分子配对，只是以尿嘧啶代替胸腺嘧啶与

腺嘌呤配对。这样，以DNA为模板，就可以合成出与其结构完全对应的RNA分子，这种RNA分子也就具有了从DNA来的遗传信息，称为信使RNA。但是作为生命活动承担者的蛋白质怎样接受遗传信息呢？它的结构与核酸完全没有相似之处。后来人们发现，在核酸中的核苷酸序列与蛋白质中的氨基酸序列之间，存在着以三个特定顺序的核苷酸与一个氨基酸相对应的关系，这就是遗传的三联密码。根据这个密码，就可以从具有特定核苷酸序列的核酸中合成出有特定氨基酸序列的蛋白质。

50年代X射线衍射DNA分子双股螺旋结构的测定，是分子生物学发展史上的一个里程碑。此后虽然又发现了DNA分子中有不同类型的双股及三股螺旋结构，但是核酸分子的主要生物功能仍然是由特定的双股螺旋承担，其他结构的功能目前还不太清楚。核酸的序列测定也是分子生物学中的一大突破，并且取得了极大的进展。全自动序列测定仪的出现和不断改进，不仅已使蛋白质氨基酸序列的测定逐渐被它的编码DNA的序列测定来代替，而且使人类基因组DNA全序列的测定得以提上日程，并引起全世界科学界的巨大兴趣及争议。人类基因组含有30亿个碱基对，它的全测定，从现在的技术发展速度来看，大约需要10～15年时间，投资30亿美元，这是生物学领域中的第一项所谓大科学。全部完成后，如果印成书，以每页3 000个印刷符号计，会有100万页，这将远远超过现有的任何一部百科全书的篇幅。美国已在国立健康研究院建立了专门的组织并着手进行，其他国家的科学界对此也极为关注。支持者认为，这一宏伟计划的完成将使我们能从根本上掌握人体所有生命活动的遗传信息，因此可以对现在还不能控制的各种危及人类生命的常见病、遗传病及癌症找到根本的解决办法。反对者则认为，人体基因组中大部分DNA序列不编码任何蛋白质，其功能还不清楚，集中大量资金解决其全序列未必能达到期望的宏伟目标，并势必影响其他急需科学项目的投资。再者，这项工作较多的是需要工作量的积累，看不出可以全面推动和促进生物学发展的前途，因此在基础研究和实际应用两方面，似乎都将是得不偿失的。

1.1.2　生命活动的基本规律

遗传信息由DNA到RNA再到蛋白质的过程是分子生物学研究的核心，通常称之为中心法则。由于两者都是由四种核苷酸（字母）组成，好

像是同一种文字的两种写法，由 DNA 到 RNA 的过程称为转录。而 RNA 到蛋白质的过程，由于两者分别由不同的字母构成（核苷酸与氨基酸），好像是两种文字，因此被称为翻译。后来发现有一些病毒的遗传信息是由 RNA 携带的，在这些病毒的复制过程中，信息由 RNA 传给 DNA，这一过程称为反转录，但这不过是极少数特例，并不影响分子生物学中心法则的普遍性。

经过多年的研究，遗传信息由 DNA 到 RNA 再到多肽链的合成过程已经基本清楚。现在的问题是，这一过程是怎样得到调节控制的。这不但是细胞发育分化的基础，也和生物体与各种环境因素的相互作用有密切关系。现在看来，调节主要发生在转录阶段，通过某些特定蛋白质与 DNA 的结合，从而控制信使 RNA 的合成。另一个问题是，蛋白质分子除一定的氨基酸顺序外，还必须有一定的空间结构才能体现生物功能。因此，怎样以一定氨基酸顺序排列的多肽链生成有一定空间结构的蛋白质，也是分子生物学中心法则目前还没有完全解决的问题。遗传信息传递的调节控制和新生肽链的折叠是目前分子生物学研究中的两个核心问题。当前，通过遗传工程或蛋白质工程得到多肽链的表达已经不会有根本上的障碍，通常遇到的困难是所得到的多肽链不能自己折叠卷曲生成有一定空间结构并具有完整生物功能的蛋白质，因此这两个问题的研究也有很大的实际应用价值。

近年来一个出人意外的发现，是某些 RNA 具有转换酶或水解酶的活性，可以把 DNA 的转录产物加工为成熟的信使 RNA。虽然 RNA 的催化活力远比一般的酶低，并且这一发现也并不影响分子生物学发展的主流，但是由于蛋白质的合成依赖于核酸的编码，而核酸的合成又依赖于酶（蛋白质）的催化，因此在生命起源问题上，长期存在着先有核酸还是先有蛋白质的疑问。RNA 既能携带遗传信息又具有酶活性的发现，为 RNA 或某些类似 RNA 的分子在生命起源过程中最先出现，而 DNA、蛋白质和酶都是 RNA 进化产物的假说，提供了有力的证据。

膜结构是生物体的基本结构之一。除细胞的外周膜之外，细胞内还有多种功能各不相同的膜结构。膜主要由磷脂和蛋白质组成，对于代谢活动旺盛的膜，如线粒体膜，蛋白质是主要成分；相反，如神经细胞的外周膜，则以磷脂为主要成分。由于磷脂是以甘油、脂肪酸酯和磷酸酯组成，因此

既有脂溶性部分，也有水溶性部分。它在水溶液中的稳定结构是以脂溶性部分内向、水溶性部分外向的双层结构，这也是膜的基本结构。蛋白质有的完全结合在膜的表面，有的穿透整个膜的双层，有的部分在膜的内部，部分暴露在表面，还有一些镶嵌在膜的内部。此外，膜结构上有时还有少量的糖和RNA。现已知道，糖与细胞间的识别及信息传递有关，它的重要性正开始受到重视。

许多极为重要的生命活动都与膜结构紧密相关。能量转换是生命活动的根本需要，无论是动物从食物的氧化中通过氧化磷酸化作用获得能量，还是植物从光能通过光合磷酸化作用取得能量都不能脱离膜的完整性。为了完成一个完整细胞的所有复杂功能，细胞内部有多种以膜包含而与细胞浆相对隔离的小体，即细胞器。每一种细胞器都有其相对独立的特定功能。例如，细胞核含有担负遗传功能的染色体，它被膜所包围，以保持染色体的相对独立性和稳定性；又如，溶酶体含有多种水解酶，仅在特定情况下发挥其功能。如果不被膜结构所包围，则将破坏细胞内部的蛋白质等重要物质，从而导致细胞死亡。

近年对于膜研究的极大重视主要来自细胞与其周围环境的相互作用。从单细胞生物来看，必须通过其外周膜才能有选择地吸取外界的营养物质，排斥外界的有害物质，使其不能进入细胞内。细胞还要通过其外周膜接受外界传来的信息，调整自己的生命活动以适应环境的变化。对多细胞的高等生物而言，除上述所有活动外，还有细胞间的物质交换和信息传递。这些重要的生命活动都要通过外周膜上的蛋白质(其中不少是糖蛋白)才能进行。细胞生物学和神经生物学领域内的许多重要问题都和膜-蛋白质体系有密切关系。

由于膜外层的亲水性质和内部的疏水性质，一般物质是不能自由通过细胞膜的，这就保护了细胞内部环境的相对稳定性，尽管外部环境处于不断变化之中，细胞活动也不受过多的影响。但是，细胞又需要不断与外部环境进行相互作用，如从外界吸取必需的营养成分，以及从外部接受信息，并作出必要的反应等。从外部吸取营养物质是通过穿透膜的蛋白质来进行的，即使是像钠离子那样的普通物质，也必需通过跨膜的有一定结构的钠离子通道蛋白才能进入细胞。实际上，对钾钠离子出入细胞的控制是神经生物学研究的一项重要内容。

细胞内外和细胞间的信息传递是通过细胞表面的一类被称为受体蛋白进行的。现在知道激素、神经递质、细胞生长因子和分化因子以及某些药物等，都首先通过它们与细胞表面的特异受体蛋白结合，才能把信息传递到细胞内部。动物体的感知行为也必需通过受体才能实现。

1.1.3 生命现象的多样性和生命基本原理的一致性

生命世界的多样性和生命本质的一致性是辩证的统一。

多少世纪以来，生物学研究的主体一直是观察和认识生命世界的多样性。从生命现象的表面观察到生命活动本质的阐明是生物学发展的必然趋势，也是现代生物学的特点。分子生物学自诞生以来短短几十年间所取得的一系列重大进展深刻地揭示了，虽然生命现象在数以百万计的不同种属中的表现形式是多种多样和千姿百态的，即使孪生兄弟也不完全相同，但是生命世界中最本质的东西，在不同生物体中却是高度一致的。

所有的生物体，从最高等最复杂的人到最低级最简单的单细胞生物，其基本组成物质都是蛋白质和核酸。它们的蛋白质都是由相同的20种氨基酸以肽键连接而成，核酸也都是同样的四种核苷酸以核苷酸链构成的。上面提到在核苷酸顺序和氨基酸顺序之间的对应关系，即遗传密码，除极少数例外，在整个生物界也是基本一致的。如果没有这种一致性，就不可能实现基因在不同生物体之间的转移及表达，也完全谈不上遗传工程和蛋白质工程。

动物和植物从外界取得能量的方式显然不同。动物从食物的氧化通过氧化磷酸化获得能量，而植物从太阳光能通过光合磷酸化取得能量。然而现在知道氧化磷酸化和光合磷酸化这两种表面看来完全不同的作用，在分子水平上的机制却是极其相似的。两者都是通过电子在一系列的蛋白质间传递，造成膜内外两侧的质子梯度差，然后合成腺三磷。整个生命世界都以腺三磷为细胞的各种活动提供能量。激素，以往一直认为是高等生物所特有的，但是近年来发现人雌性激素受体在酵母中表达后，对酵母细胞基因的转录可以进行依赖于雌性激素的调节。哺乳动物胰岛细胞分泌的胰岛素也被发现在某些单细胞生物中存在，并表现和在高等动物中类似的促进生长的功能。生命活动高度一致性的阐明说明分子生物学确实已经深入到了生命现象的核心和本质。

1.1.4 生物学的全新面貌

当前凡是研究生命现象的学科，不可避免地都要深入到分子水平去进行本质规律的探讨，这使分子生物学很快就渗入生物学的各个领域，改变了整个生物学的面貌；同时也对医学和农业科学及其应用产生了巨大的影响。生物学的全新面貌最突出地表现在出现了一系列新的分支学科，如：分子遗传学，分子细胞学，分子分类学，分子神经解剖学，分子药理学，分子病理学，分子流行病学等等，影响到生命科学的所有领域，即使是生态学、古生物学和分类学等也不例外。下面简略地介绍分子生物学对一些基础生物学领域的影响，着重当代生物学研究的几个热点，以及虽然从表面看来和数理学科关系较远，但由于分子生物学的冲击已成为面貌一新的学科。

作为生命体基本单位的细胞和作为生命活动最高形式的神经活动，是现代生物学研究最活跃的领域。今天，细胞生物学和神经生物学由于采用了分子生物学的新思想和新手段而获得了新的生命力，研究步伐大大加快，从而与分子生物学一起发展成为当代生物学研究的三大热点。现在如果浏览一下国际上这两门学科的主要刊物，或者重要国际会议的内容，就可以知道，其内容已经和传统的细胞生物学及神经生物学的内容大不相同，而是以分子生物学为基础，在分子水平上进行研究，实际上已经演变成为分子细胞生物学和分子神经生物学。

细胞学的基本知识早就告诉我们，所有细胞的分裂都有一个周期。直到现在，我们才从分子水平认识到细胞是如何控制其分裂周期的。一种称之为周期蛋白的物质在细胞内的浓度发生着周期性的变化，它的合成使其浓度不断增加，在细胞分裂之前达到高峰。周期蛋白活化了细胞内的蛋白激酶，后者的活力升高引起一系列化学变化，最终导致细胞分裂。在细胞分裂时，周期蛋白则迅速被水解，浓度下降。在子细胞中周期蛋白又重新合成并积累，开始下一个周期。控制这个周期的分子机制，在从酵母到人的细胞中是完全一致的，这再次说明生命活动在分子水平上的高度一致性。

一个单一的受精卵细胞怎样在发育过程中，逐渐分化成为一个由千千万万不同类型和不同功能的细胞组成的成熟个体，一直是细胞学研究的核心问题之一，也是发育生物学的中心内容。现在知道，细胞的分化也

是由基因决定的，并处于由基因编码决定细胞分化期间的蛋白质的控制之下。例如，编码控制果蝇复眼细胞分化的有关蛋白质的基因已被分离，DNA 序列测定的结果表明，它和癌基因有类似的特点，所编码的蛋白质是一个膜蛋白，与某些生长因子受体结构相似，很可能具有酪氨酸蛋白激酶的活性。细胞发育分化的研究实际上已经是基因调控的分子生物学研究，而基因调控的分子生物学研究必须上升到去解释整个细胞活动的水平。各个学科之间的渗透和交叉如此密切是现代生物学的又一特点。

除物质交换外，细胞与外界环境以及细胞间的信息传递依赖于存在于细胞表面的或跨膜的蛋白质。这些蛋白质称为受体蛋白，它们接受外界信息，然后把信息传达到细胞内部，再引起一系列的变化，调节细胞的新陈代谢，以适应外部的环境。这些蛋白质在生命活动中有多方面的重要作用。例如，某些生长因子就是通过受体蛋白控制细胞的分裂繁殖。这种受体蛋白能与鸟三磷(GTP)结合，因此被称为 G 蛋白。这个 G 蛋白-GTP 复合物能够把细胞生长的信息传递到细胞内，但它又有 GTP 水解酶活性，能在以上过程中把鸟三磷水解为鸟二磷(GDP)。所生成的受体蛋白-GDP 复合物则是细胞生长终止的信号。现在知道，G 蛋白在细胞生长调节等方面有重要作用。癌细胞生长失调也与此有关。

在神经生物学领域中，神经细胞中信息的存储和组织，以及神经细胞间信息传递的研究，都已进入了分子水平。就电生理学而言，脑组织中的电流活动与膜的离子通道密切相关。这些离子通道都是跨膜蛋白，通过蛋白构象变化控制离子出入神经细胞。这些离子通道蛋白大体上可分为配体控制(如乙酰胆碱受体)和电压控制(如钠离子通道蛋白)两大类，它们都具有特定的专一性。现在不少通道蛋白的编码基因 DNA 序列，及由此得到的蛋白质序列已经测定，在膜上的结构与其作为离子通道的功能关系也已有所了解。

动物一般有极强的感觉能力，这是动物生存所不可缺少的，也是神经生物学研究的重要内容之一。某些动物感知气味的灵敏度及分辨能力都是极其惊人的。灵敏度可达一万亿分之一，并能分辨化学物质的光学异构体。现在这些现象已经从分子水平开始得到阐明。视觉、味觉和嗅觉等都有自己的受体，这些受体也都是蛋白质。这些受体蛋白质不但有极高的专一性，并且以极高的亲和力与配体结合，这就是生物体感觉的分辨力和灵

敏度的分子基础。

大脑的研究也许是神经生物学领域中最复杂的问题了。思维、感情、学习、记忆都是大脑的活动，不少人认为它将是21世纪最活跃的研究领域，现在也已开始进入分子生物学研究的范围。笔者并不怀疑21世纪大脑研究将有突破性的进展，但最主要的进展必将从分子水平上取得。人的大脑可能含有100万亿个突触接点，其物质基础仍然离不开受体蛋白。正是这些突触接点的连接组织方式，决定了人脑的活动。记忆的基础可能与突触接点连接组织方式的改变有关。某些蛋白质的磷酸化能对突触接点的加强或减弱起调节作用，从而影响突触的连接组织方式。

分类和进化是生物学领域中历史最悠久的学科之一。近年来受分子生物学的影响，这两门古老的学科已经产生了根本性的变化。过去主要是依靠生物体的形态，辅以生理特征探讨亲缘关系的远近，进行分类和研究不同种属间的进化关系。对于形体简单的生物体，如细菌等，生理特征的比较就成了主要手段。近年来由于蛋白质和核酸序列研究的突破性进展，人们逐渐发现了不同生物体中担负相同功能的蛋白质和核酸在序列上的相似性和它们在亲缘关系上的远近有直接的联系。由于生物体适应环境的能力，其形态和生理受它们所生活的环境的影响很大，而蛋白质和核酸序列则反映了不同生物体进化演变更为根本的性质，较少受环境影响，因此也更为可靠。

目前应用较多的是核糖体RNA序列的比较。在原核生物中，过去的习惯分类法是分为蓝绿藻和细菌。现在根据RNA的序列，更为合理的分类是分为古细菌和真细菌。前者包括甲烷菌、硫细菌和嗜盐菌；后者则包括其他较为常见的细菌如杆菌等。两者在细胞膜脂质的化学组成上也有明显的不同。关于真核生物的分类学，人们利用核糖体18 S RNA序列的比较，也解决了一些过去不清楚的问题。

除直接的核酸或蛋白质的序列比较外，还可以用较为简单的不同种属间单股DNA的杂化，或抗原抗体的相互作用来分别比较DNA或蛋白质结构的相似性。根据碱基配对原理，不同种属间DNA序列越相似，其单股DNA配对后生成的双股DNA分子就越稳定。抗原抗体的相互作用也有类似的关系。这些方法都是不需要复杂的技术或设备就可以做到的。对于高等动物所得结果虽然大体上和经典的看法近似，但也澄清了一些

过去有一定争议的问题。例如，根据DNA的杂化和免疫蛋白的相互作用，明确了熊猫属于熊科，而不是属于猫科等。和传统的方法相比，这类方法的另一个优点是可以根据核酸或蛋白质序列差异的程度，判断种属之间的分化时间。例如，熊猫与熊科其他种属分化的时间是1 500～2 500万年前，人类和黑猩猩的分化时间是在630～770万年前。

由于冷冻的尸体、骨骼甚至化石中的DNA分子仍能保存一定的完整性，这就使对现已灭绝的生物种属在分子水平上研究其进化和分类成为可能。例如，西伯利亚冻土中保存的猛犸象尸体中的DNA仍然相当完整，其序列分析结果充分说明猛犸象和现代象有密切关系。已知保存最久的DNA或许是保存于琥珀中的象鼻虫(黎巴嫩)的DNA，已有1.2亿年，但序列仍然相当完整，从其序列可以确定其与现有种属的关系。在美国爱达荷州湖底发现的叶化石中的DNA也有1 700万年以上的历史，它是编码光合作用中一种极为重要的酶——核酮糖-1,5-二磷酸羧化酶加氧酶的基因片断，从DNA序列比较使对这种现已灭绝的植物与现存种属的关系的研究成为可能。类似的研究已形成了一门新的学科，即分子古生物学。

考古学从表面看来和生物学关系不大，甚至可以被认为不属于自然科学。但是考古学家现在也不能不对分子生物学有所了解。近年来对保存了几千年的埃及木乃伊进行研究的结果，发现其体内的DNA分子仍然保存了一定程度的核苷酸顺序完整性，由于人体内的所谓卫星DNA的核苷酸顺序是极端多变的，每一个人都不一样，不同个体之间DNA顺序的相似性依赖于他们之间的血缘关系。因此对埃及王族木乃伊的DNA的序列分析，可以对埃及王族的世系有所阐明。

1.1.5 分子生物学对人类生活的巨大影响

分子生物学的兴起还不到40年，它所取得的成果，已经在工业、农业及医药卫生等方面有了重要应用。特别是基因在不同个体、甚至在不同生物种属之间的转移，为某些遗传疾病的治疗和新品种的培育等提供了前所未有的可能性，从而为人类健康、农业增产、以及控制和改造整个地球上的生物界展现了无限广阔的美好前景。但从近年来的发展看，这还仅仅是一个开端，将来必然会有更为广阔的发展前景。分子生物学成果在工业上的应用产生了以基因工程为基础的生产生物制品的新兴的工业。从一

种生物体分离编码某个蛋白质的基因，即 DNA 片断，再人工重组到可以用发酵法大量生产的如大肠杆菌或酵母的基因中去，使其在大肠杆菌或酵母的细胞中表达成这个蛋白质，并达到大量生产的目的。新近发展起来的蛋白质工程则是分离出某个蛋白质的基因之后，根据三联体密码，把这个 DNA 序列中编码某一个氨基酸的密码子，改变成为编码另一个所需要的氨基酸的密码子；或者用合成 DNA 的方法直接合成基因，在体外加以改造。通过以上两种方法都可以得到在天然界原来并不存在的 DNA，再用类似上述的方法，引入到大肠杆菌或酵母的基因中进行表达，以达到大量生产的目的，得到具有新特性的蛋白质。

组织纤溶酶原激活物(t－PA)是一种有助于溶化血栓的蛋白质，在生物体中含量甚微，不可能用天然来源来制备药物。现在已经用基因工程方法大量生产，并被用于中风的预防和治疗。仅此一种产品，年产值已达 2.3 亿美元。糖尿病是一种常见病，过去治疗糖尿病的胰岛素的生产依靠从动物胰脏中提取，经常供不应求。目前世界市场上销售的胰岛素，用基因工程方法生产的已占总产量的一半，为糖尿病患者带来了福音，年销售额约为 5.7 亿美元。估计到 1997 年，生物高技术产品的世界总销售额将达 71 亿美元。

基因工程用于农业，已经对农作物的品种改良起到了前所未有的重要影响。农作物以及家畜品种的改良，现在可以用定向引入有关基因的方法进行，这就从根本上改变了过去盲目大量诱变然后再从中进行筛选的传统做法。在农作物中，已经成功地对马铃薯进行了改造，不但使其获得了抗病毒基因，也得到了高蛋白质含量的马铃薯新品种。把一个蛋白质水解酶抑制剂基因引入烟草之后，使得以烟叶为食的害虫不能消化其中的蛋白质而不能繁殖，这样，这一烟草品种就获得了抗虫害的能力。对番茄的基因改造，得到了比较不易软化和擦伤的品种，因此可以在成熟后收获并且保存较长时间，也避免了过去在成熟前收获因而口味不好的缺点，此产品已经在美国上市。虽然植物基因工程的应用还刚刚开始，但为农作物的大量增产和品种改造，例如固氮基因的转移等，提供了无法估量的发展前景。

癌症是危害人类最严重的疾病之一。分子生物学的进展为认识癌症发病原因，从而彻底征服癌症提供了新的可能性。现在已经发现了几十种

癌基因,其中已有十余种被认为与癌症病源学有关。前面提到生长因子受体蛋白,即G蛋白,能与GTP及GDP生成复合物,从而实现对细胞生长的控制。某些癌基因的表达产物,p21蛋白与G蛋白的氨基酸序列极为相似,但p21蛋白与GTP生成复合物后,却不能引起GTP水解为GDP从而终止细胞的分裂,这就导致了细胞生长的失控。最近,p21蛋白的X射线衍射空间结构已经解出,对它和GTP及GDP的结合方式,包括分子中与结合有关的肽段,及直接与GDP结合的氨基酸残基,都已有所了解,并发现它分别和GTP及GDP结合后,分子的空间结构发生明显的变化。最近,还利用同步加速器的强X射线光源,对p21蛋白水解GTP的动态过程进行了研究。这些都将对癌症的最终解决有所帮助。

人类的遗传疾病是由于基因缺陷而引起的。早已知道镰刀状贫血症是由于血红蛋白基因中的一个核苷酸的突变造成蛋白质中一个氨基酸的突变,从而引起脱氧血红蛋白溶解度下降,在细胞内成胶或聚合,使红细胞变形成为镰刀状,并且丧失结合氧分子的能力。另外一种比较常见的遗传疾病是由于编码苯丙氨酸羟化酶的基因丢失,人体不能合成苯丙氨酸羟化酶,造成苯丙氨酸在体内的积累而引起痴呆症。此外,某些过去曾被认为是生活习惯的问题,现在也证实与遗传有关,并发现其分子水平上的原因。酗酒基因的发现及克隆就是近来一个引人注意的例子。用引入有关基因的方法,来治疗遗传疾病已经取得了很大的进展。在一些国家基因治疗已经取得初步成果。看来,在不远的将来,许多遗传疾病可望得到治疗。

人体内的所谓卫星DNA序列是极端多变的,对每一个人都不一样。现代技术的发展,只要极微量的材料,如一滴血,一根头发,一块皮肤屑或极少量的精液,从中提取的DNA经过多聚酶链反应扩增,就可以对卫星DNA进行序列测定。罪犯可以极小心地不留下指纹,但很难完全避免留下身上极少量的上述物质,这一方法已经被用于刑事案件中罪犯的确定,被称为DNA指纹法。又由于不同个体之间DNA序列的相似性是依赖于他们之间血缘关系的远近,根据卫星DNA之间序列相似性的比较,不仅可以决定亲子关系,甚至可以区别亲兄弟和堂兄弟,这在民事案件中也十分有用。DNA指纹法已经开始在实际案件中普遍使用,所得结果发生错误的可能性仅为一亿分之一。和任何新方法一样,使用不当也会发生错

误，但随着方法的不断完善，可能会取代指纹法成为法医学中的重要方法。

分子生物学的发展已经揭示了生命本质的高度一致性，这是人类认识自然和认识自己的极大飞跃，下一世纪的生物学将是对生命活动本质有统一认识的真正的统一生物学，21 世纪将是统一生物学的世纪，并将形成崭新的生命观。分子生物学对科学和人类生活的全面影响，完全可以和本世纪初物理学所引起的变革相比，它和生物化学及生物物理学一起，吸引了大量的物理学家和化学家进入生物学领域，从而也影响了这两个学科的发展。它不仅现在是自然科学中发展最迅速的领域，到 21 世纪，它仍将毫无疑问地是自然科学领域中的带头学科。

参 考 文 献

[1] 邹承鲁. 结构生物学的时代已开始. 科技导报，1995，**4**：7
[2] Edgington T S. *FASEBJ*, 1994，**8**：1213
[3] Commentary. *Nature*, 1993，**362**：783
[4] *Science*，1994，**263**：319
[5] Sternberg M J E，*et al*. *Trans R Soc Lond*[*Biol*]，1994，**344**：365
[6] Hillis D M，*et al*. *Science*，1994，**264**：671
[7] Finkel T，Epstein SE. *FASEB J*，1995，**9**：843

（邹承鲁）

§1.2 关于理论物理与理论生命科学的一些思考

1.2.1 普遍与特殊

物理科学和生命科学的相互作用由来已久，电学和电生理学是同时诞生的（L. Galvani，1791）。生物学为物理学启示了能量守恒定律，物理学为生命科学提供了从显微镜、X 光，到示踪原子、中子衍射、核磁共振谱仪、同步回旋辐射、扫描隧道显微镜等种种研究手段。然而，只有到了把生命物质和生命过程作为真正的物理对象加以研究的时候，理论物理的思维方法和数学工具才会发挥更大作用。现在，生命科学所积累的大量事实和数据，物理科学在研究复杂系统方面的长足进展，已经把物理学和生命

科学密切结合的任务提上日程。对于理论物理工作者而言，恰当地认识生物学和物理学思维方法的差异，会更有助于寻求共同语言。

物理学研究物质结构和运动的普遍规律，这些规律适用于人类观测所及的宇宙的任何部分，适用于它的过去与未来。物理规律可以由少量关键实验证实，物理规律没有例外。观测与理论的矛盾，通常表明理论还不够普遍，预示着理论的突破，而新的理论总把旧理论的正确部分作为特例包含在内。物理理论从数学框架到实验检验两方面受到严格制约，很难出现学派林立、假说纷纭的局面。物理规律本身是否也在演化，在更大程度上是个哲学问题，至少其变化的时间尺度远大于现代人类的观测时间，因而在当前的认识水平上，“永恒”的物理定律被认为是放之宇宙而皆准的。

生命科学研究的对象极为特殊，迄今还没有发现任何其他天体上有类似地球上的生命现象。就宇宙尺度而言，我们目前所研究的对象，只存在于微乎其微的小小地球的“生物圈”的有限空间，存在于狭窄的温度范围，存在于出现生命后的有限的历史时期。为特殊现象建立理论是更困难的，但其内容却可能极其丰富。生命科学的理论和规律常有例外，它们既不能还原成简单的物理法则，也不会违反基本的物理规律。

1.2.2 遍历与进化

物理学处理“多体”“多自由度”，甚至“无穷多自由度”的问题时，标准方法是统计平均。把对时间、对历史、对发展过程的平均，换成对一切可能的当前状态的平均。这是基于对经典力学的遍历假设，即认为一个力学系统或迟或早会经历能量上允许的一切状态。遍历假设早在物理学范围之内就出了问题：原来认为应当遍历的许多体系，例如耦合非简谐振子，已被严格证明并非如此。不过，确实还有许多遍历性成立的情况，例如耗散系统的混沌吸引子上的运动。

与遍历类似但更弱一些的要求，是时间变化的定态假设，即认为一个发展过程总处在同一机制、同一环境之中。其实，定态过程是无始无终、“始终如一”的。遍历和定态都是与进化或演化对立的概念。

虽然我们现在看到的宇宙，也处于演化之中，而且对种种遥远距离的观测带来不同演化时期的信息，但那时间尺度比之人类历史是极其漫长的。因此，物理学家们往往从现存的结构和相互作用来解释现象，更习惯于在“遍历”和“定态”的假设下思考问题。

进化是生命发展的基本特征，生命现象只有在进化背景中才能正确理解。进化乃是一种“遍历破缺”，这“破缺”在生物学中叫做“选择”。

另一方面，相对于单个生物体甚至物种的寿命，生命进化过程也是极其漫长的。大自然有足够的时间来“择优”。彭桓武教授曾经引用老子的话“天地不仁，以万物为刍狗”，来概括生物进化过程的漫长和物竞天择的“无情”。

1.2.3 因果论和目的论

在生命科学文献中，拟人的、“目的论”的语言颇为常见，而因果论则是物理学的一条基本原则。从表面看，这两者似乎矛盾。细想起来，当物理学处理比较复杂的系统时，已经有转向“目的论”说法的迹象。举两个例子。

描述布朗运动的朗之万方程中有摩擦系数和随机力强度两个参量，两者不能任意给定，而必须满足涨落耗散定理。这是为了保证时间趋向无穷时布朗粒子要和周围环境达到热平衡，使得能量均分定理成立。实质上，这是从终值而非初值条件选取随机微分方程的解。统计物理学中许多关于趋近平衡条件的讨论都有这么一点“目的论”的味道。其实，趋近平衡只是最简单的一种非平衡过程。处理远离平衡现象时，可能更要注意“目的”，才能正确描述。

自旋玻璃的基态和蛋白质折叠问题，都涉及在高维的能量空间中寻求最小值。这是多参数的优化问题，其提法本身就是要求“目标函数”达到极值。一般说来，优化问题常与一定的识别问题相联系，但往往更困难。穷举一切可能的状态，有点“遍历”的精神，但不能解决规模较大的实际问题。近年发展起来的某些有效的近似方法，例如“模拟退火”，实质是选取合理有利的演化途径，尽量接近某种事先设定的目标。演化的每一步，应遵从因果律，但长期演化所达到的状态，看起来像是有目的安排的结果。

目标函数的极值也有绝对、相对之差。没有理由认为复杂系统的演化一定沿最优方向、发展到最优状态。有许多足够好的演化方式，可以达到相当“优”的状态，后者在实践上与最优状态相差不远。不同演化道路的初等实例是优选法：按 0.618 作“黄金分割”，比使用 0.5 的“对分法”，在有限精度下快不了几步。复杂性不是随机性，但复杂性自然地包含着随机性，在“同样”复杂的条件下，一个系统如果有两次演化机会，不一定走同

一条演化道路。

1.2.4 特殊中的普遍

生命现象涉及众多的层次,每个层次上都有空间结构和时间进化,层次之间并没有简单的"自相似"。理论物理工作者比较宜于从具有普遍性,可以"言之及物"的那些层次入手,争取逐步由"物"及"生"。

十多年来,我国已有不少理论物理工作者在各种层次从事研究,例如:生物大分子的结构;它们之间、它们和中小分子以及环境(例如水)的相互作用;核酸碱基序列、蛋白质氨基酸序列的信息结构分析;蛋白质的变性、复性、折叠和高级结构;生物膜的连续介质模型,特别是液晶模型,膜的静态形状和动态行为,膜上的和跨膜的物理过程。

在组织、器官、个体乃至群体、种群、生态系统的诸层次上,"物"理逐渐成为并不直接考虑的因素。许多唯象的模型研究,更近乎数学生物学。细胞内和细胞间的通讯、正常和病态的心律、脑电波分析、神经脉冲传递、免疫系统等方面的研究,不少可以借鉴非线性科学的方程和语言,同熟知的数学物理模型相联系,都还是理论物理工作者的用武之地。这里要注意,普适性的理论,如相变、混沌、分形等,不能代替对特殊问题的具体研究。用新颖的名词术语包装,成不了科学成果。

有一些永恒的问题,如生命的起源和本质,意识的产生和发展,人类必将世世代代思索和探求。对生命物质的结构、运动和相互作用已经作过系统研究的人,或许能对这类根本性问题悟出些深刻的道理。但这些问题看来不宜作为研究的起点。

在若干年内,我国的计算机仍将同美、日等国有相当差距。我们不能也不应在拼机器能力的"从头计算"或大规模的分子动力学模拟方面与这些国家抗衡,而应当注重"巧算",发展粗粒化和唯象的方法。用化学反应动力学语言,平衡和非平衡热力学概念与工具进行的研究,也属于粗粒化和唯象的理论,也应提倡。不过,下面只讨论这几年用粗粒化方法研究非线性动力学[1]所引发的一些思考。

1.2.5 粗粒化和符号描述

自然科学不能同时在一切层次和细节上把握客观现象的规律性,而必须聚焦到一定的观测精度和描述水平。当我们写下质子和中子的符号p和n时,通常不必关心它们由哪些夸克组成,而只须知道其电荷和质

量。在化学中使用原子符号写出分子结构式时，H、C、O、N 等记号主要代表一定的原子量、离子半径、亲和力、化学价，而无须涉及原子核的构造。对于由几十个原子组成的嘌呤和嘧啶，化学结构式并非处处必需。人们引入 A、C、G、T 等等符号，更注意它们形成双螺旋时提供两个还是三个氢键。20 种符号所代表的不同氨基酸，更重要的不是"大同"的化学结构，而在于"小异"所导致的亲水、疏水、极性等种种性质的差别。

粗粒化不可避免地导致符号描述。在核酸和蛋白质的情形下，粗粒化都导致了一维的符号序列。其实，对于有限长的符号序列，一维和高维并没有原则性的差异。高维序列可以归结为具有远邻关系的一维序列。这当然比只有最近邻关系的一维序列复杂些。但对于多数一维序列，也必须研究其中的长程关联。人们用符号序列来表示研究成果，更希望从符号序列中解读出自然界的规律。

面对大量符号序列，首先会想到统计分析，例如计算和比较各种符号串的出现频度与关联。事实上，人们用统计方法对核酸和蛋白质序列作了大量研究。罗辽复教授 1995 年 4 月在国家自然科学基金委员会召开的一次座谈会上曾说，多年来用统计方法分析核酸编码序列的经验表明，这些序列的统计性质往往偏离随机序列不远。这是否表示，传统的统计分析不足以把生命现象远离平衡的"远"充分放大，以供更细致的分类和刻划？这也提出了对符号序列进行确定论分析的必要性。其实，对于有限长的序列，确定论分析和统计分析都是近似的，但恰当的粗粒化可能导致某些严格的结论。

1.2.6 形式语言和语法复杂性

从一定的字母集中选取符号，形成一维的符号序列。这些符号序列的集合，构成某种形式语言。粗粒化和符号描述自然地导致形式语言。对形式语言和自然语言都可以进行统计分析。然而，形式语言也把我们带到因计算机科学发展而成熟起来的另一个领域，开辟了对符号序列进行确定论分析的可能性。

在 50 年代中期，Chomsky[2]就发展了一种形式语言的分类体系。他把语言分成四大类，每类语言可由一类计算机接受（或者说识别、执行）：由不需存储器的有限自动机接受的正规语言，由带堆栈的下推自动机接受的上下文无关语言，由存储器比例于输入量的计算机接受的上下文有

关语言，由具备无穷存储的图灵机接受的递归可数语言。这同时给出复杂性的一种参考尺度，通常称为 Chomsky 阶梯。

Chomsky 的阶梯，并不是对形式语言的唯一的分类体系。在符号动力学所导致的符号序列的分类上，已经感到其不足。例如，谢惠民等人的研究表明，一维单峰映射中的符号序列，从简单的正规语言一跃而至上下文有关语言，跨越了上下文无关语言的整个台阶。在图象表示领域，也发现许多直观上极其简单的图象，如直线段或圆，竟然导致复杂的上下文有关语言，使得研究上下文无关语言所得到的丰富成果无用武之地。

1968 年荷兰生物学家 A. Lindenmayer 在研究多细胞植物的发育过程时，建议了一种符号描述方式。它后来发展成另一套形式语言的分类体系，现在称为 L 系统。Chomsky 用串行的生成规则。L 系统用并行生成，它所导致的分类比 Chomsky 更细致，但看来并未覆盖一切可能的形式语言。Chomsky 分类中的某些复杂序列，在 Lindenmayer 系统中处于简单层次，反之亦有实例。符号动力学中郑伟谋建议的广义生成法则，自然地对应 L 系统的框架。L 系统的一大优点，是它与分形几何有自然的联系。用基于 L 系统的图形生成程序，模拟草本植物的宏观形态发育过程，已经达到可以乱真的程度。对于核酸碱基序列或蛋白质氨基酸序列的分析，L 系统或其变种，可能发挥作用。

参考文献

[1] 郑伟谋，郝柏林. 实用符号动力学. 上海：上海科技教育出版社，1994
[2] 谢惠民. 复杂性与动力系统. 上海：上海科技教育出版社，1994

（郝柏林）

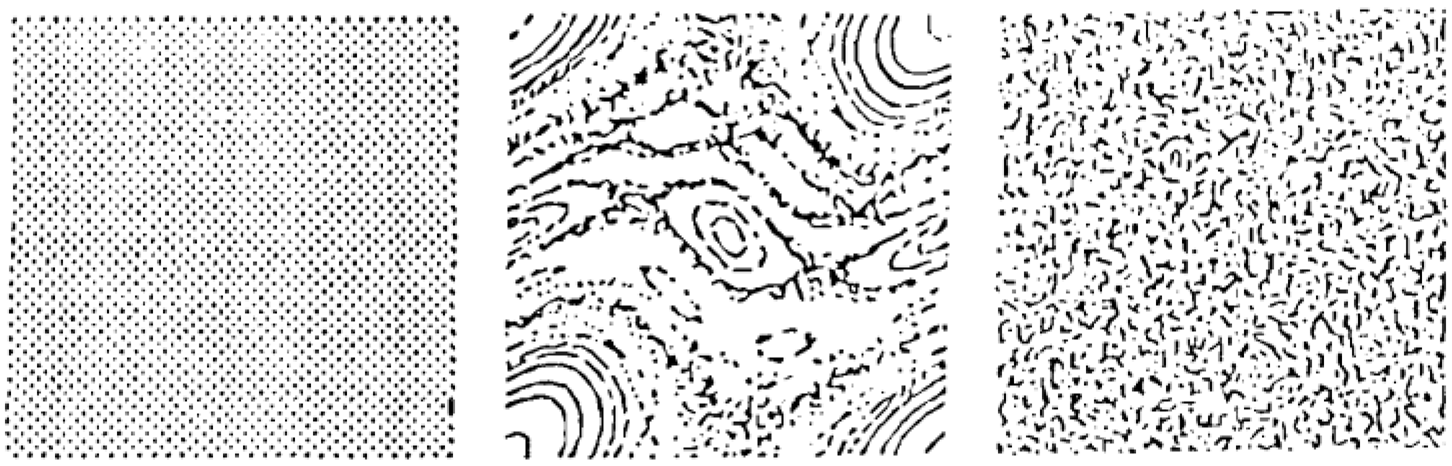
§1.3 复杂性与生命现象

1.3.1 复杂性

人所共知，生命现象是复杂的。但是复杂性作为一个物理的量度是个新问题。自 40 年代末 Shannon 提出信息论以来，生物学家都认为：我们

有了一个可以帮助研究复杂问题的量度了。但当时有人认为信息量并不能作为复杂性唯一的量度，这两个概念之间还有区别。如 von Neumann 就指出“解释复杂和复杂性概念应该是 20 世纪科学的任务，就像 19 世纪的熵和能量概念一样。”

大家知道信息的概念是相对的，只有在传输过程中才能得到具体而确定的值。Shannon 的信息量只代表了事物的随机性，而没有描述事物的规律性。复杂性应该包含有规律性和随机性的因素，Grassberger[1]有一张图很能说明这个含义（图 1.3.1）。



(A) (B) (C)

图 1.3.1 复杂性意义示意图

(A) 有规律图象；(B) 复杂图象；(C) 随机图象。

从这张图来看，左边的图是有规律的图象，右边的是随机的图象，而只有当中的图是个复杂的图象。但是如果我们认为这类系统可以用百分之多少是随机成分、百分之多少是规律的成分来表达那就错了，因为通过滤波把这两者分开是线性系统的特征，线性系统中复杂性的问题不是主要的，而复杂性的问题出现在非线性系统的发展过程中。

在还没有清楚地认识复杂性问题的本质时，人们企图多方面来刻画复杂性。下面举个例子。

冗长度

重复(repetition)

对称(symmetry)

均匀(homogeneity)

简并(degeneracy)

无特殊性(unspecific)

可交换(interchangeable)

时间上重复(repetition in time)

→因果性，决定性

(causality，determinism)

复杂性
变化(variety)
非对称(asymmetry)
非均匀(heterogeneity)
特殊(specific)
独特的(individuality)
出于意料的(unexpectedness)
→时间上新出现的,不确定性,随机性(newness in time, uncertainty, randomness)

随着非线性科学的发展,人们注意到它的意义在于从根本上动摇过去科学的框架。过去人们认为科学可以帮助人们预见事物将来的发展,就像19世纪牛顿力学所取得的辉煌成果那样。但是世界终究是"复杂的",像牛顿力学所能描述的现象并不多,绝大部分是不能用力学的原理去解决的,于是物理学家就发展了一套统计的方法来处理随机现象。但自从混沌运动被人们知道以后,决定性与随机性的划分已经不清楚了。为了刻划事物的特征,复杂性的问题就成了当前的热点。

最初也许是Kolmogorov提出所谓算法复杂性(algorithm complexity)[2]具体概念的量度,以他的说法,对一件事物的描述所用的最短的计算机语言的长度可以用来衡量复杂性。但是这种量度本身仍然与熵是等价的,那就是说衡量了事物的随机性。

以后又有人如Wolfram[3]从元胞自动机(cellular automata)研究中发现复杂性可分等级,如平凡的吸引子、周期运动、混沌和更为复杂的现象。这些发展又与形式语言和自动机联系起来成为一套与符号动力学相关的理论体系。关于这方面的发展现在已有苏州大学的谢惠民教授写的一本书《复杂性与动力系统》[4],在这里就不多讲了,但是从中可以看到复杂性问题的研究不简单。为了跟上现在的发展,这里还可介绍一篇文献[5],其中有各种复杂性的定义,以及用这些定义所导出的算法用来测试抛物线映射(logistic map)的结果。

我们从直觉出发总以为任何复杂性的量度应该是简单和普适的,要不然自己就会陷入复杂的陷阱中。虽然上面讲过很多复杂的特征,但我们估计复杂性的测度应该包含有两个基本的因素,即规律或结构性和随机性,再有就是它们之间的关系。从这种直观的观点出发我们曾定义了两种测度[6],称之为C1和C2复杂度,一个代表了随机性,与KC相等价,一个是衡量结构性的。我们从研究脑电图时间序列大脑皮层的信息传输中,得到一些有趣的结果。

在这里我们可以举出一些研究脑电图时间序列复杂性的工作，从而可以看到复杂性的测度对研究生物学问题的重要性。

人脑的活动对环境的变化是非常敏感的，因此要研究脑的活动除去用分析的方法，即现在神经生理学的主流以外，还必须在完全不(或最少)干扰的情况下来研究它的活动。现有的方法并不多，如正电子图象扫描、NMR 造影、脑血流图等。但这些大都以脑的生物化学反应为依据，因此都只能反应脑中的大的、慢的变化。脑磁图和脑电图却是能够及时反应脑的功能变化而又对受试对象干扰较小的方法。但脑磁图的设备昂贵，比较实用的还是脑电图。不论脑电或脑磁的变化都非常没有规律，用混沌动力学理论还不能满足脑电的复杂性的要求，所以我们采用 Vastino 和 Swinney[7]的互信息理论来研究多导脑电图之间的信息传输。例如在八导脑电图的记录上可以得到相互之间的信息传输时间序列，这八导的记录位于 F3、F4、C3、C4、T3、T4、01 和 02，这是以国际标准命名的，可简称为 1、2、3、4、5、6、7 和 8(图 1.3.2)。

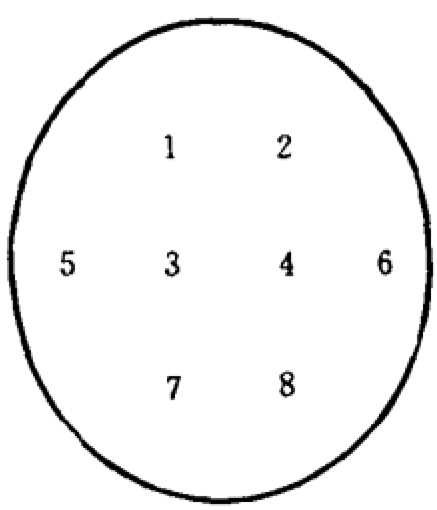


图 1.3.2 脑电图电极位置示意图

用时间延迟的方法可以得到信息传输的时间序列如图 1.3.3，这是一个青年人的 2 导至 5 导的记录。采样频率为 1 000 Hz，时间为 768 ms。这种时间序列用复杂性的 Kolmogorov 量度 KC[8]及 C1 和 C2 测量就可得到复杂度的数值。如果同时测量，就得到 8×8＝64 个数值，把它们排成矩阵，那么矩阵的行为从该处输出的信息，列为该处收到的信息。图 1.3.4为对一个人在进行算术运算时的记录。从该图可以看到第五导处(左颞叶)最为活跃，它发出的与收到的信息复杂度都高于其他部位。用这种方法我们可以直接观察到人脑的功能活动[9]。

看来复杂性的问题还远未解决，对今后生命科学的发展有着重大的意义。这里谈的都只涉及到一维的主要的复杂性问题，如时间序列。作为生命现象来说，显然要讨论时空的复杂性问题。现在可能有个比较近似的时空复杂性研究的想法，即简单地从一维的基础上推广。研究一维我们用的是矢量距离，如果用矩阵间距离的话，可能适用于某些生命现象，因为由此可引入对称性破缺的考虑。

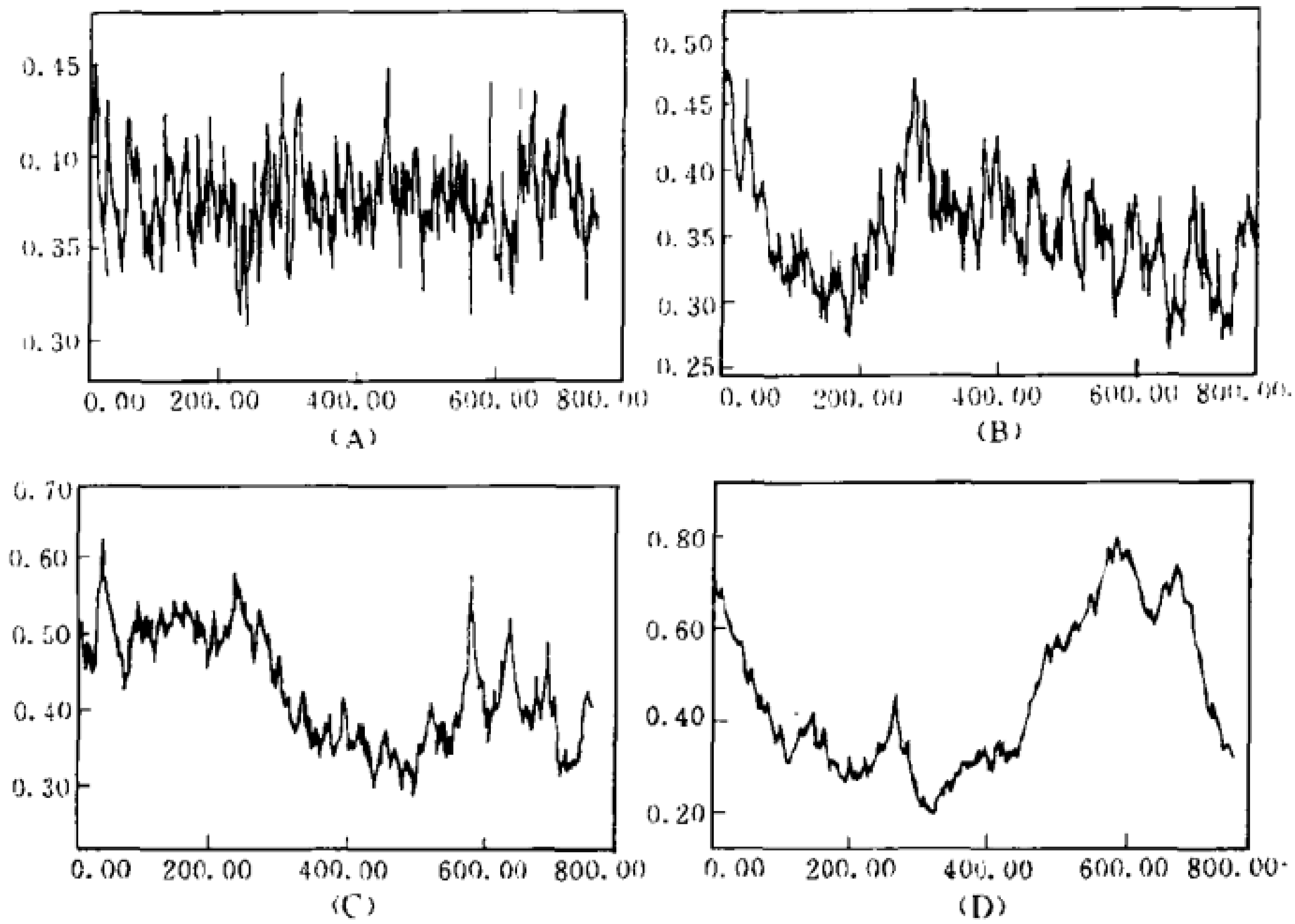


图 1.3.3 用时间延迟方法得到的信息传输时间序列

(A) 正常睁眼；(B) 闭眼；(C) 轻度睡眠；(D) 深度睡眠。

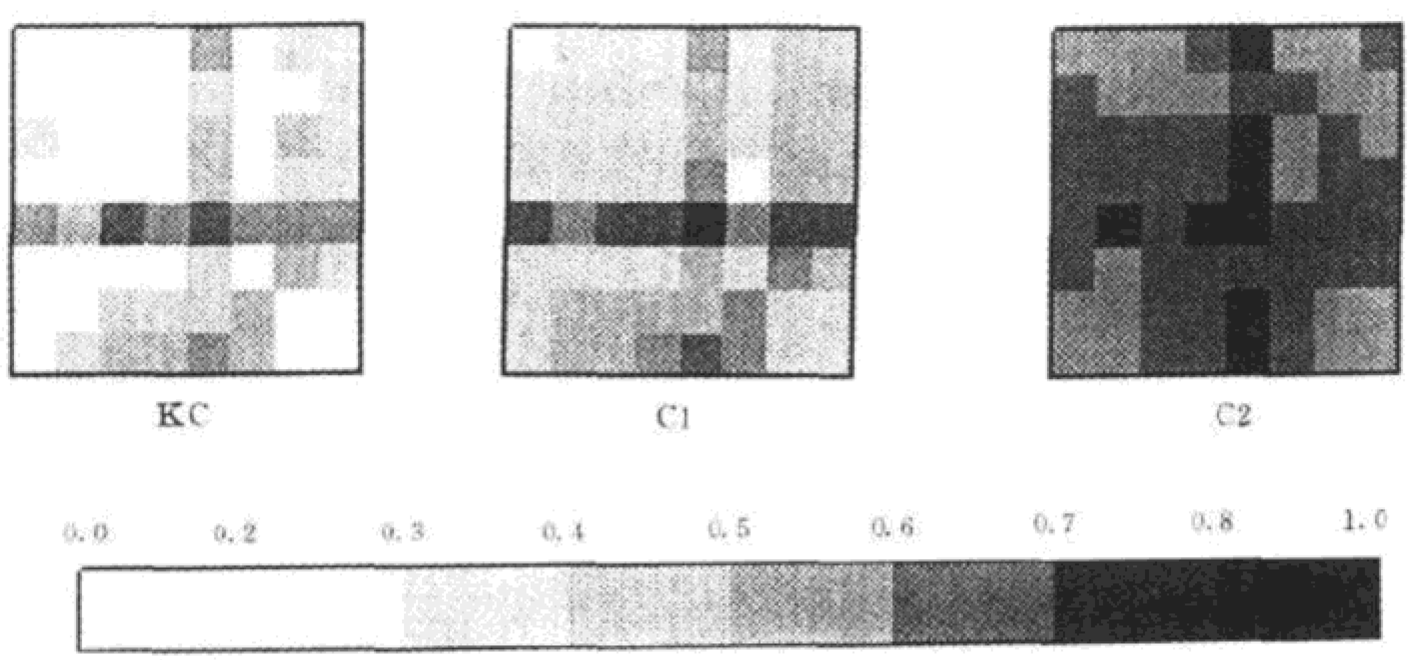


图 1.3.4 大脑皮层信息传输矩阵

1.3.2 生命现象的根本问题

我们知道生命现象的复杂性不单是个量度问题，它有着深刻的含义和长久的历史。

40 年代，一位英国的理论生物学家 Woodger 写了一本书，名叫《生

物学原理》(*The Biological Principles*),这本书中谈到了生物学的四大矛盾,即活力论与机械论(vitalism vs. mechanism),新生论与预成论(epigenesis vs. preformation),因果论与目的论(causality vs. teleology),个体与环境(organism vs. environments)。半个世纪过去了,我们现在回头再来看一看这些问题究竟解决了多少。这 50 多年生物学有了很大的发展,这些进展是如此巨大,以至于把前面的问题都推到后台去了。大家都忙于搞甚么基因工程、生物医学工程、蛋白质工程等等,对刚才说的问题大概都认为是老夫子的谈话,但实质上这些问题仍然存在并未解决。一些有深刻思想的生物学家,如分子生物学家 Jacob Monod 在去世前曾对此做过严肃的思考,他写过《可能性与必然性》(*Chance and Necessity*)等等的书和论文。特别在当前讲究经济效益的时候,谈论这类基本理论问题是极不适宜的,但是如果从它们的重要性来看却是很根本的。这类争论或矛盾经过多年的讨论和研究可以说有了一些眉目,这就是它们都与生物学与物理学理论基础的统一有关。既然我们要讨论生物学与物理学的关系,那么只好认命吧,来讨论一下这些老掉牙的又没有解决的老问题。这些问题分析起来可以说又都与热力学第二定律有关。

从 40 年代薛定谔写了《生命是什么?》(*What is life?*)[10]一书以后,经过 Turing[11]、Chandraskhar[12]到 Prigogine[13],又过了差不多 20 多年的时间,我们才知道在远离平衡态时某些系统会自发地产生结构,也就是说有这样的系统可以把自身产生的熵排出到周围环境中去,从而减少自身的熵,使其熵低于周围环境。这样做并不违反第二定律,但是,这是一种"自私"的行为。譬如在夏天装了空调自己舒服了而使得别人更热,就是这个原理的简单的表现。现在我们可以详细地描述如何制造一个系统使自身的熵降低。但是如何自发地出现这样系统的普适规律却不清楚。这个问题很重要,因为它与生命起源有关。我们从上述的研究结果已经知道,在一个动物系统中出现结构常常是由于系统的某些动力学变量发生了对称性的破缺的缘故。生命系统中有一特殊的破缺,那就是分子手征破缺。

在现实中,如果我们要探索宇宙间其他地方有无生命,唯一指标就是用有无分子手征性破缺来判断。在本世纪的 70 年代,美国成功地登上了月球,一时间探索所谓球外生物就成了热门的话题,甚至有了一本杂志叫《球外生物学》(*Exobiology*)。但经过一个时期的探索,基本上否定了有球

外生物的存在，起码在我们可及的宇宙中没有任何迹象表明有生命物质。这个工作主要是由美国航天局的研究机构进行的。他们的方法就是用从球外取得的样品和天外飞来的陨石分析，看有无手征破缺的分子。他们的研究结果否定了这类物质的存在。过去很多肯定的报道被证明是由于陨石进入大气层后被手征破缺物质污染了。

那么我们就要问：地球上的手征破缺的分子是怎么产生的呢？人们发现所有的生物体内的主要分子都是手征破缺的。这个发现就是巴斯德在1845年观察到的生物体的分子光学活性问题。直至现在我们对这个现象的基本原理知之甚少，但是这个指标却是我们用来探测球外物体有无生命的唯一的可靠的标准。所以生物体的分子手征性的破缺是个大问题。

生物体分子手征性的问题有两个方面。第一方面是破缺的起源。巴斯德早年曾经说过：不对称的结果必然由不对称的原因引起。所以在很长的一段时期中有很多工作者从事于寻找分子手征不对称的原因，多少年过去了，没有很好的结果。后来宇称不守恒的原理被发现了，于是乎又热闹起来，但是后来证明也是很困难的，因为宇称不对称的β粒子的能量与分子的能量级相差约有10个数量级。虽然到现在人们还没有放弃这个问题的探索，但就笔者看来，这个问题相当困难，困难不在于发生破缺，因为我们知道宇宙间有很多破缺，如物质、宇称和时间方向等等都是破缺的。困难在于绝对对称的状态是非常不稳定的，从非线性科学的发展中我们知道微小的扰动会在临界点或分岔时起着决定性的作用。而要去追究亿万年前的这种小扰动如何发生，看来是很困难的。第二个方面是这样的系统有哪些特性使得它与生命现象有如此紧密的关系。这个问题却是一个可以而且能够从理论物理学的角度来讨论的。我们知道生物分子的手征特征在于它们的旋光性，即对平面偏振光的偏振角旋转了一个角度。根据经典的或量子的电磁学，我们可以用如下的方程来描述这个现象：[14]

$$\overline{\langle\mu\rangle_{nn}}=\alpha_n E-\gamma_B B+\frac{\beta_n}{c}\frac{\partial B}{\partial t},$$

$$\overline{\langle m\rangle_{nn}}=\kappa_n B+\gamma_n E+\frac{\beta_n}{c}\frac{\partial E}{\partial t}。$$

这里$\langle\mu\rangle$是分子的电偶极矩，$\langle m\rangle$是分子的磁极矩，α、β、γ、κ都是一些系数，它们都有对称性的约束。由于分子内部的不对称性使得电子在分子内

运动是弯曲的路线和不平均的分布，由此产生电偶极矩和磁极矩。手征分子的特点就在于它可以有电磁耦合的性质。

根据统计力学的理论，任何动力学的变量相互间产生耦合的话，它们之间的对称性必须要匹配。譬如说，一个时间不对称的量$\langle A\rangle$与一个空间不对称的量$\langle B\rangle$，它们之间的相关函数$\langle AB\rangle$必然要等于零。因为在时空反演算子作用下，这个相关函数变成自身的负值。而在手征破缺的系统中，不同对称性的时空变量是可以耦合的。在手征系统中这种反演使之变为它们的镜象，所以手征分子对动力学变量的时空耦合有选择性。这对生物学来说有很大的意义。因为生物学的语言中常常使用的就是结构与功能，用信息论的语言则是时间信息与空间信息的转换。在生物复杂的动力系统中，如果它既不是随机的也不是完全有规律的，而是“复杂”的，在其运动发生分岔时手征选择的考虑应是个重要的问题。[15,16]

我们知道，过去热力学第二定律被认为是经验的，由于宇宙学的发展把第二定律的起因拿去了，这使我们松了一口气。负熵概念的提出，耗散结构的理论的建立，使我们又松动了一下。但是如何自发地形成一个局部熵降低的系统，即能够把自身的熵排出到体外，这仍然是个大问题。看来这种熵的不对称的分布也很可能与分子手征不对称有关。

说来说去还是那个老问题，生机论与机械论的争论。说穿了生机论与机械论的差别只在于机械论认为生命有如机械，生机论认为生命是有如机械，不过要有个驾驶员。如何去找驾驶员本身就是寻找生命的本质。不同的时代有不同的机械，从钟表、热机、电机到计算机都成为说明生命的例子。直到现在神经网络的发展致使 Penrose 写了一本叫《皇帝的新脑》的书，他在这本书中提出了“我”在哪里的问题?实质上也就是驾驶员的问题。

参 考 文 献

[1] Grassberger P. *Inter J Theor Phys*, 1986, **25**: 907
[2] Kolmogorov A N. *Inf Trans*, 1965, **1**: 3
[3] Wolfram S Commun. *Math Phys*, 1984, **96**:15; *Phys Rev Lett*, 1985, **55**: 449
[4] 谢惠民. 复杂性与动力系统. 上海:上海科技教育出版社, 1994
[5] Wackerbauer R, *et al*. *Chaos, Solitons and Fractals*, 1994, **4**: 133

[6] Xu J H, Liu Z R, Liu R. *Chaos, Solitons and Fractals*, 1994, **4**: 2111
[7] Vastino J A, Swinney H L. *Phy Rev Lett*, 1988, **60**: 1773
[8] Kasper K, Schuster H G. *Phys Rev*, 1987, **A36**: 842
[9] Xu J H, Wu X B. *Science in China*, 1994, Ser B **37**: 1455
[10] Schrödinger E. What is Life? Cambridge University Press, 1961
[11] Turing A M. Phil Trans Roy, London: Ser B, 1952, **37**: 237
[12] Chandraskhar S. Hydrodynamic and Hydromagnetic Stabilitiy, 1961
[13] Nicolis G, Prigogine I. Self-orgnaization in Nonequilibrium Systems. John Wiley & Sons, 1977
[14] Caldwell D S, Eyring H. The Theory of Optical Activity. John Wiley & Sons, 1971
[15] Xu Jinghua, Ding Da fu. *Scientia Sinica*, 1981, **24**: 997
[16] Xu Jinghua, Ding Da fu. *Scientia Sinica*, 1983, **28**: 250

(徐京华)

第 2 章　蛋白质与生物大分子

§2.1　蛋白质分子运动

现代生物学的发展与物理学家的参与是分不开的。仅就近年而言，新兴的结构生物学就是物理学家把对晶体结构的研究方法应用到生物大分子（主要是蛋白质及核酸）三维结构解析的结果。分子生物学也是由于物理学家克里克和生物学家沃森合作研究提出 DNA 的双螺旋结构模型而发展起来的。物理学家 Gamov 提出三联体学说，开创了遗传密码的研究。近年来晶体学家重视 DNA 与蛋白质相互作用的研究，使 DNA 的复制、转录、翻译和调控等分子生物学的重要机制获得了合理的解释。

由于大量的蛋白质三维结构研究成果的积累，现在已有可能从蛋白质的静态研究进入动态的分析。近年的研究表明，蛋白质分子是不断运动的，分子内部结构包括结构域、α 螺旋等都可以运动，从而执行各种功能。整个蛋白质分子如肌球蛋白的运动可以产生力，使肌纤丝收缩或微丝运动。下面讨论一下近年来这方面的一些进展。

2.1.1　蛋白质的结构生物学

自 1912 年 Max von Laue 发现 X 射线及 1913 年 Bragg 提出 Bragg 方程后，晶体学得到了快速发展。Bernal 和 Crowfoot（Hodgkin）率先研究蛋白质的晶体结构，他们于 1939 年对胃蛋白酶进行了晶体衍射研究。在此以前德国著名化学家 Willstatter 认为蛋白质是胶体，是一种无序的结构。Bernal 等证明胃蛋白酶并非胶体，而是一种有序的三维结构[1]。Perutz 等于 1952 年后首先研究了血红蛋白和肌红蛋白的晶体结构，Kendrew 于 1957 年解出了肌红蛋白的高分辨结构，Perutz 于 1959 年解出了血红蛋白的高分辨结构，两人共享诺贝尔奖。Hodgkin 等于 1972 年测出胰岛素的晶体结构，我国梁栋材也得到了胰岛素的高分辨结构。自 1972 年以后，溶菌酶、胰蛋白酶、胃蛋白酶等大批蛋白质晶体结构研究得到

发展，到了1980年以后蛋白质晶体结构的研究成果直线上升，1990年测出了110余种蛋白质，目前已达到平均每天完成一种蛋白质测定的速度。

由于晶体学技术的快速发展，现已广泛采用同步辐射X射线作为光源，面探测器收集衍射数据以及计算机图形工作站等手段，在许多亚细胞如核糖体及各种病毒的结构方面取得重要进展。此外结构生物学在蛋白质与蛋白质相互作用（如抗原抗体关系的研究）及蛋白质与DNA相互作用（如拓扑异构酶I）等研究都获得重要成果。更有意义的是近年来利用晶体学在研究蛋白质分子的动态上取得了突破，为今后生物学与物理学的结合开辟了一条途径。

最近在晶体学技术方面也有很大发展，普林斯顿大学物理系与康乃尔大学合作，将电荷耦合装置CCD引入到面探测器中，使得分辨率大大提高（<0.1 nm），如伴刀豆球蛋白A的结构由原来的0.16 nm分辨率提高到0.098 nm，从而可以从电子密度图直接观察到原子的位置[2]。

2.1.2　蛋白质分子的运动

利用X射线衍射所测定的蛋白质晶体结构虽然是静态的，但是利用不同条件下所培养的晶体研究构象变化，或者用不同生物材料所获得的同种蛋白质的晶体研究其构象，就可能得到蛋白质结构的动态变化。近年来有些学者还用Laue衍射研究蛋白质的动态结构。

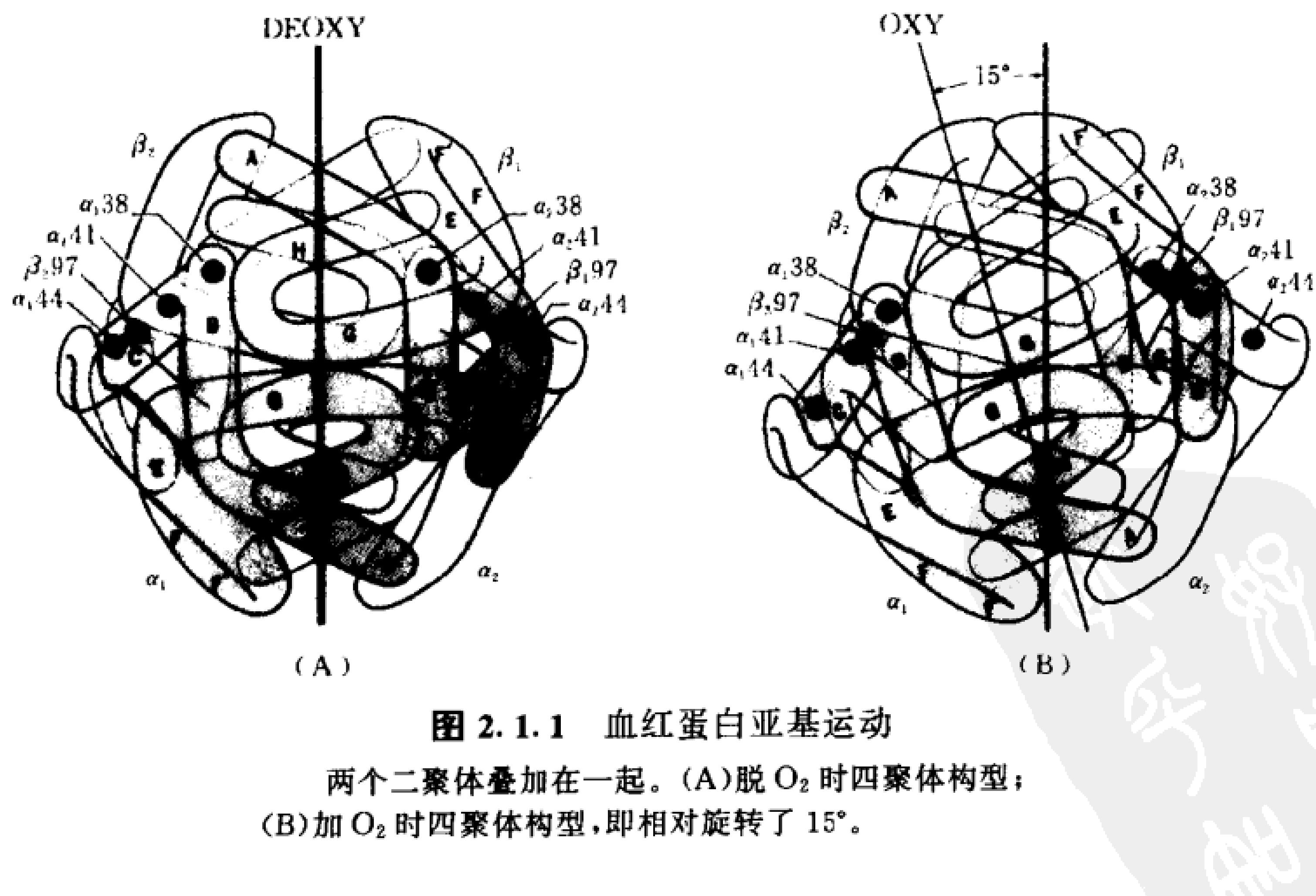


图 2.1.1　血红蛋白亚基运动

两个二聚体叠加在一起。(A)脱O_2时四聚体构型；(B)加O_2时四聚体构型，即相对旋转了15°。

(1) 血红蛋白的分子运动

早在1970年Perutz即研究血红蛋白四聚体与O_2结合后的构象变化，他发现$\alpha_1\beta_1$二聚体与$\alpha_2\beta_2$二聚体可以相互运动。Perutz最初测定血红蛋白结构时是测定脱氧血红蛋白在无氧条件下培养出的血红蛋白晶体。以后他在有氧条件下培养出氧合血红蛋白的晶体，测定这个晶体结构时他发现血红蛋白与O_2结合后其构象发生了改变。氧合血红蛋白与脱氧血红蛋白的空间结构不一样。血红蛋白四聚体由两个相同的二聚体($\alpha_1\beta_1$及$\alpha_2\beta_2$)组成，在二聚体中亚基紧密结合。在氧合血红蛋白与脱氧血红蛋白中α、β链的三级结构并无不同，但是血红蛋白与O_2结合后，它的四级结构却发生了变化。血红蛋白的两个二聚体能够彼此相对移动，氧合血红蛋白分子更加紧密，"$\alpha_1\beta_1$对"改变它与"$\alpha_2\beta_2$对"的相对位置，转动的角度达到15°，这种变化由图2.1.1可以看出。当血红蛋白失去O_2时，它又缓慢地移动，重新回到原来的构象。血红蛋白与O_2的这种相互运动，犹如肺的呼吸，因此有人称之为分子肺。

最近Jayaraman等用共振拉曼光谱研究血红蛋白与CO结合形成的HbCO，当照光后HbCO解离时中间结构所发生的变化。这些光谱变化揭

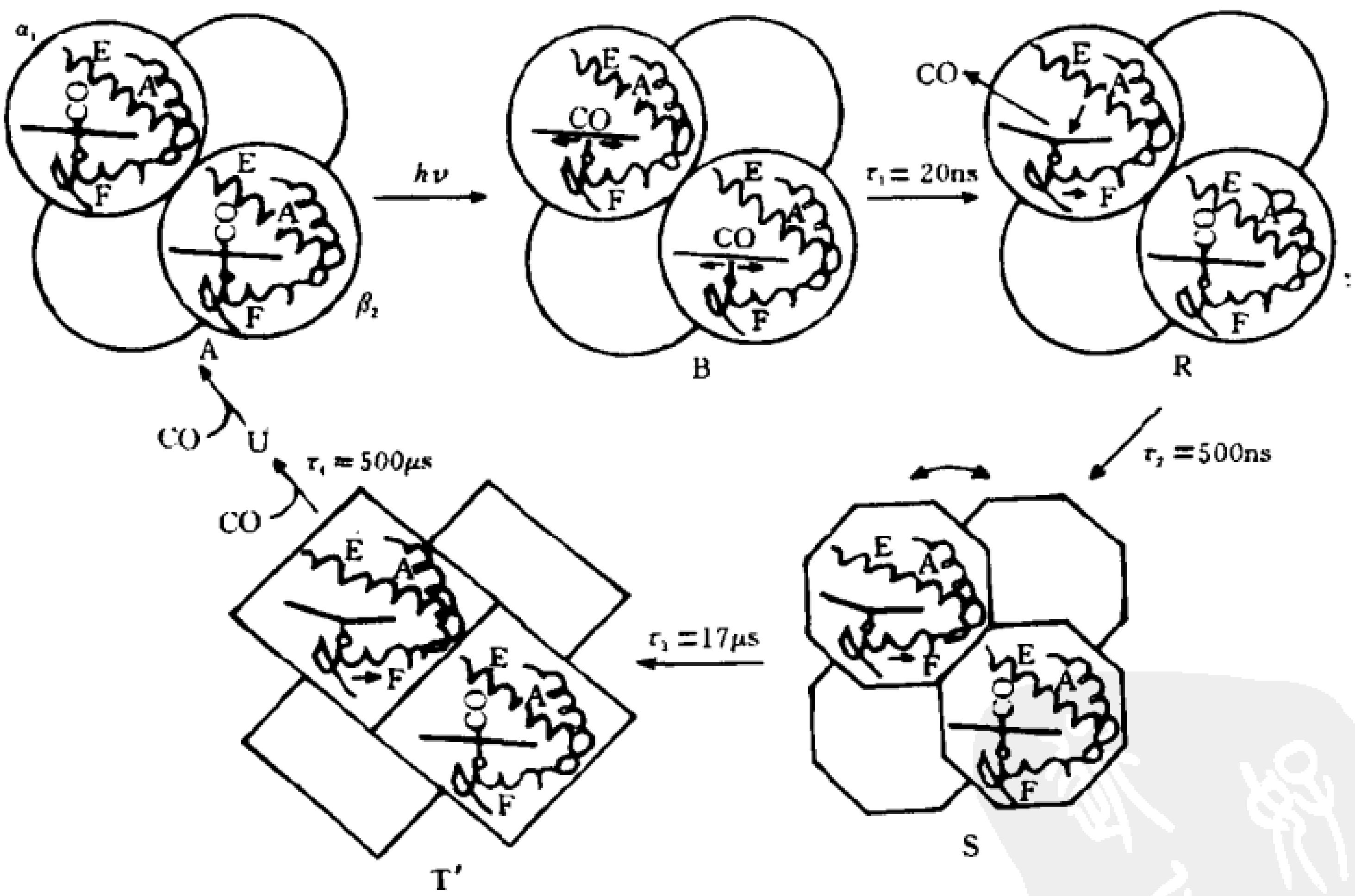


图2.1.2　血红蛋白反应的模型

HbCOA的光解产生耦合态B，然后产生中间态S。四聚体中的箭头表示运动。

示了照光过程中血红蛋白的三级结构及四级结构的一系列运动，这些运动包含有远端及近端螺旋和 $\alpha_1\beta_2$ 亚基界面的运动[4]，见图 2.1.2。

（2）腺苷酸激酶的分子运动

酶分子进行催化反应时它的分子构象不断地发生变化才能完成催化作用。Schulz 于 1974 年即测定出腺苷酸激酶的晶体结构。腺苷酸激酶催化腺苷酸(AMP)的磷酸化，使 AMP 生成腺苷二磷酸(ADP)，同时 ATP 也生成一分子 ADP，反应如下：

$$AMP+ATP \rightleftharpoons ADP+ADP$$

Schulz 等一直研究腺苷酸激酶分子的动态变化，他认为腺苷酸激酶与底物的关系符合 Koshland 的诱导契合学说(induced-fit-theory)[5]。

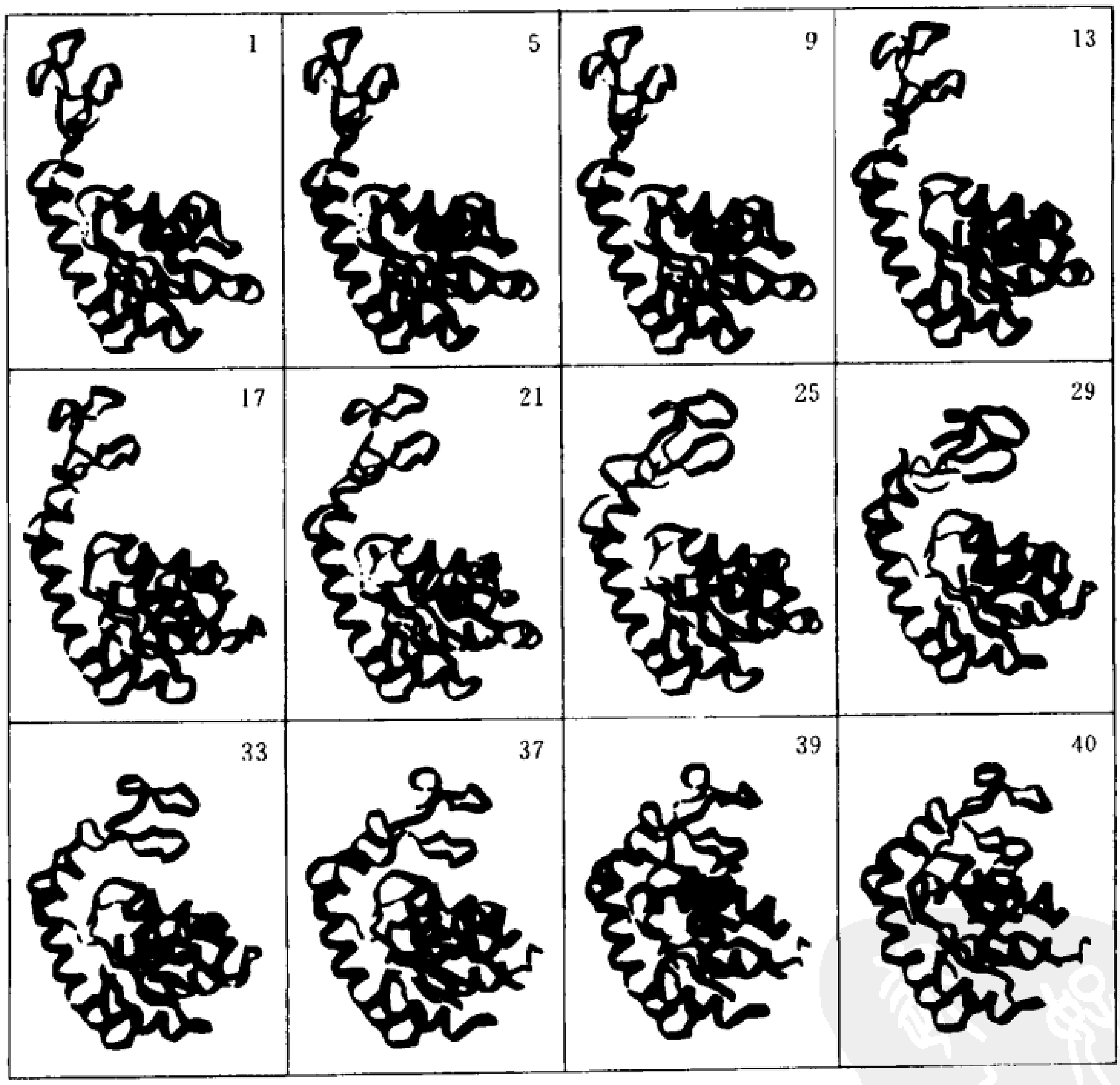


图 2.1.3 腺苷酸激酶动画电影的一部分

表示腺苷酸激酶分子的动态变化。

Schulz 等已将 17 种腺苷酸激酶的晶体结构解析出来，包括大肠杆菌、酵母、牛、猪、鲤鱼等。他们选出 9 种结构，分辨率在 0.163～0.26 nm 范围之内，按照其结构的重叠情况，将它们排列起来。腺苷酸激酶与底物(AMP)结合有两种构象，一种为开放态构象，另一种为闭合态构象，还有一些为中间过渡态。他们以腺苷酸激酶的核心结构域的 β 折叠作为基准，将所有中间过渡态都排列起来，并用计算机将所缺少的中间态加以补充，将这些结构连接起来形成一个光滑运动，这样显示出来就成为一段电影。他们将腺苷酸激酶分子运动制成了录象带，放映出来就显示出酶分子在催化过程中的运动状态，见图 2.1.3[6]。

关于酶分子运动的研究还有很多，如柠檬酸合成酶等，在此不一一列举。

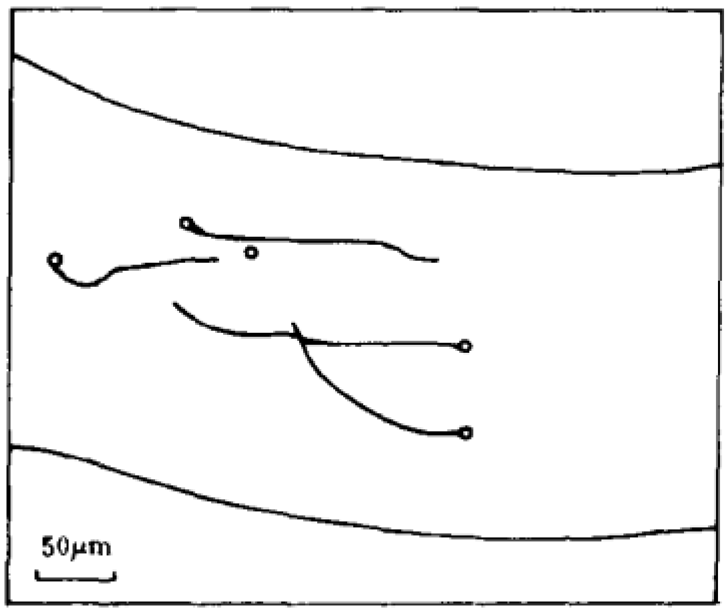
(3) 肌肉收缩与细胞运动

肌肉收缩和细胞运动都是依赖蛋白质的运动完成的。Huxley 于 1969 年提出肌纤丝滑动学说解释肌肉收缩的机理，他认为肌肉收缩时肌纤丝长度的变化是由于粗纤丝(由肌球蛋白组成)与细纤丝(由肌动蛋白组成)彼此相互滑动的结果。他的理论得到大量实验证明。肌动蛋白的晶体结构已于 1990 年由 Kabsch 等解出，肌动蛋白能聚合成细纤丝[7]。肌球蛋白头部的三维结构于 1993 年由 Rayment 等测定出来[8]。根据晶体结构研究结果，Rayment 提出了肌肉收缩的分子模型，基本上与 Huxley 的模型相似[9]。

Spudich 等用一种细胞很大的丽藻(*Nitella*)作材料研究细胞运动，丽藻细胞中含有规则排列的肌动蛋白丝。他们将用肌球蛋白的头部(即重酶解肌球蛋白，HMM)包被的能发荧光的塑料小球(直径 0.7 μm)注射到丽藻细胞内，在荧光显微镜下可以观察到塑料小球沿着由肌动蛋白组成的缆绳向前运动，并可以拍摄照片。这是由于塑料小球表面上的肌球蛋白头部与肌动蛋白丝相互作用的结果[10]。

刘雄等用电视显微摄象和计算机图象处理技术观察了月见草(*Oenothera*)花粉管中细胞器沿肌动蛋白丝束的运动，结果与 Spudich 看到的结果相似，可以观察到细胞器沿微丝束在运动。如果用细胞松弛素 B (cytochalasin B，一种使肌动蛋白丝解聚的试剂)处理，则细胞器运动立即停止，只作布朗运动。前人已证明在细胞器表面上存在着肌球蛋白，由此可见细胞器靠肌球蛋白与肌动蛋白丝相互作用而产生运动[11]。

笔者用显微录象技术观察了烟草花粉管及水霉(*Saprolegnia*)菌丝内的颗粒运动,细胞颗粒沿着由肌动蛋白丝组成的轨道作定向运动[12,13]。图 2.1.4 为月见草花粉管中细胞器运动的轨迹。



2.1.4 月见草花粉管内细胞器运动的轨迹

柳田敏雄等将肌球蛋白分子均匀地涂布在玻片上,用荧光标记肌动蛋白并聚合成肌动蛋白丝,置于玻片上,加 ATP 后在荧光显微镜下可以观察到肌动蛋白丝在肌球蛋白推动下的运动,并进行了录象[14]。

最近 Spudich 小组和柳田敏雄小组分别用光钳技术将一条肌动蛋白丝固定住,用少量重酶解肌球蛋白分子结合在发荧光的塑料小球上,使一个肌球蛋白分子头部与肌动蛋白丝接触。他们用灵敏的光电二极管检测装置可以检测出单个肌球蛋白分子所产生的力。单个肌球蛋白分子平均产生的力为 3～4 pN,一个肌球蛋白分子推动肌动蛋白丝的位移为 11 nm。这是迄今为止单个肌球蛋白分子与单个悬浮的肌动蛋白丝相互作用所产生的力和位移的直接测量结果[15,16],见图 2.1.5。这些结果是肌肉收缩或细胞运动时,肌动蛋白与肌球蛋白相互作用将 ATP 的化学能转换成机械能的充分证明。

2.1.3 展望

从上面的几个例子可以看出,在生物细胞中运动是经常、普遍发生的,不但可以观察整个细胞的运动,如原生动物、单细胞绿藻和细菌的运动,而且在细胞内各种细胞器和颗粒也不断地进行运动,甚至蛋白质和酶分子也都时刻地进行运动。运动是生物有机体的基本性质和功能。

蛋白质分子运动的研究受到物理学家的高度重视,例如 B. B. Levi

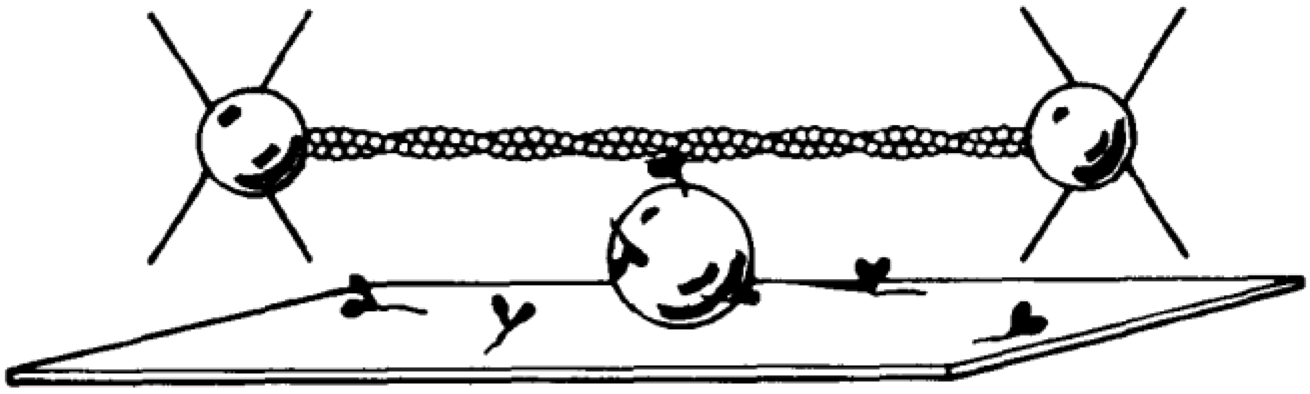


图 2.1.5 肌球蛋白分子与肌动蛋白丝的相互作用

肌动蛋白丝的两端用光钳固定在塑料小球上。包被在中央玻璃小球上的一个肌球蛋白分子位于肌动蛋白处。肌动蛋白丝一端的位移用光学检测。(引自文献[15])

最近在 *Physics Today* 上发表文章,讨论了肌球蛋白和驱动蛋白马达分子驱动微丝和微管的运动[17]。

结构生物学不但可以提供蛋白质的空间结构,而且从蛋白质结构的动态研究还能揭示蛋白质分子的性质和功能,如构象的变化、电子的转移等,这些资料对生物大分子理论物理的研究将会有重要的参考意义。相信生物大分子的理论物理学将会使生物学的研究进入更深的层次,进一步揭示出一些生命问题的本质。

参考文献

[1] Perutz M. Protein structure: new approaches to disease and therapy freeman. New York, 1992

[2] Waoker R, Gruner S M, Ealick S E. High resolution macromolecular structure determination using CCD detectors and synchrotron radiation. *Structure*, 1995, **3**: 835

[3] Dickerson R E, Geis I. Hemoglobin: structure, function, evolution and pthology. Benjamin, 1983

[4] Jayaraman V, Rodgers K R, Mukerji I, *et al*. Hemoglobin allostery: resonance Raman spectroxcopy of kinetic intermidiates. *Science*, 1995, **269**: 1843

[5] Schulz G E, *et al*. Induced-fit movements in adenylate kinase. *J Mol Biol*, 1990, **213**: 627

[6] Vonrhein C, Schlauderer G I, Schulz G E. Movie of the structural changes during a catalytic cycle of nucleoside monophasphate kinase. *Structure*, 1995, **3**: 483

[7] Kabsch W, Mannherz H C, *et al*. Suct D Atomic stucture of the actin:DNaseI complex. *Nature*, 1990, **347**: 37

[8] Rayment I, Rypniewski W R, Schmidt-Base K. Three-dimensional structure of myosin substragment-1:a molecular motor. *Science*, 1993, **261**: 50

[9] Rayment I. The three-dimensional structure of a molecular motor. *Trends Bioch Sci*,

1994, **3**: 129

[10] Sheetz M P, Spudich J A. Movement of myosin-coated fluorescent beads on actin cable *in vitro*. *Nature*, 1983, **303**: 31

[11] 刘雄,阎隆飞. 月见草花粉萌发后花粉管和花粉粒内细胞的运动. 植物学报,1993,**35**:859

[12] 阎隆飞,唐孝威,刘国琴. 花粉管胞质颗粒运输的流动轨道系统. 自然科学进展,1994,**4**: 599

[13] 刘雄,阎隆飞. 水霉(*Saprolegnia*)菌丝内细胞颗粒运动的特征。植物生理学报,1993,**21**: 131

[14] Harada Y, Yanagida T. Mechanochemical coupling in actomyosin energy transduction studied by in vitro movement assey. *J Mol Biol*, 1990, **216**: 49

[15] Finer J T, Simmons R M, Spudich J A. Single myosin molecule mechanics: piconewton forces and nanometre steps. *Nature*, 1994, **368**: 113

[16] Ighijima A, Harada Y, Kojima T. Single molecular analysis of the actomyosin motor using nanomanipulation. *Biochem Biophys Res Comun*, 1994, **199**: 1057

[17] Levi B G. Measured steps advance the understanding of molecular motor. *Physics Today*, 1995, **4**: 17

(阎隆飞)

§ 2.2 蛋白质中的自折叠单元

蛋白质的折叠和去折叠是现代结构生物学中的一个基本课题。研究蛋白质折叠的最终目的是为了解开氨基酸序列是怎样决定结构的这个谜,这将使我们能够从一个蛋白质的序列出发,预测其三维结构,也将使我们对于蛋白质结构的形成过程及稳定性等有更加深入的理解。

将一个蛋白质分解成几个较小的组成部分,如果这些部分仍然保留有折叠的能力,并且折叠成与这些部分在整体蛋白中一样的结构,这部分结构就称之为"自折叠单元"。自折叠单元可以被看成为构筑一个较大的复杂的蛋白质的基本结构因子。它的序列中仍包含决定它自己结构的充分和必要的信息。

研究蛋白质折叠的最大障碍之一是它们具有很大的自由度。一个典型的蛋白质是一个由几百个乃至上千个氨基酸残基组成的非周期性长链分子。它所能够形成的可能构象的数目是一个天文数字。显然,一个蛋白质不可能通过遍历所有的可能构象来找到它的最佳结构。如果能将一个蛋白质分解成若干个由几十个氨基酸残基所组成的自折叠单元,那我们就能够大大简化对蛋白质折叠的研究。现有的计算方法已有可能预测一

个由几十个氨基酸残基组成的多肽(小蛋白)的结构[1,2]。

对一个已知结构的蛋白质,试图找出一种合理的分割方法,来预测其中哪些部分可以形成自折叠单元。先作如下的假设:①分割时不切断蛋白质中的二级结构单元(α 螺旋或 β 折叠),因为蛋白质中的大部分二级结构单元均由 10 个左右的氨基酸残基组成,被切断的小片断很难有形成同样结构的可能。②能够形成自折叠单元的那些部分相当于密致排列的二级结构单元的集合,它们之间的相互作用应超过它们与蛋白质其他部分的相互作用。这些假设正确与否,将通过比较预测与实验的结果来加以验证。

如果把一个蛋白质的结构用一个数学上的"图"来表示。图上的每一个顶点表示一个二级结构,而每一条边则表示两个二级结构单元之间的相互作用。那末,这些相互作用的强度将通过每两个二级结构单元间有多少对重原子(指碳、氮、氧等)之间的距离在 0.5 nm 之内来决定。两个相邻的或平行的二级结构单元之间的相互作用将大于距离较远的或垂直的二级结构之间的相互作用。

然后,我们将对这个图作如下的简化:我们将只保留对于每一个顶点最强的相互作用及超过这个最大值 60%的那些相互作用。对每一个顶点来说,它的邻点对它的相互作用密度定义为它与邻点的相互作用除以代表这个顶点的二级结构的长度。我们也将保留相互作用密度超过整个图中最强相互作用密度 20%的那一部分,其余的相互作用所对应的那些边将被舍弃。这样所得到的图将是非对称的,即对某一顶点来说,它的邻点对它来说可能是重要的,而同时这个顶点对它的邻点来说却是不重要的,因为它的邻点与其他顶点有更强的相互作用。见图 2.2.1(B)和(C)。

在经过简化的图中,寻找具有自折叠能力的部分相当于寻找这个图的最大连通子图(maximal connected subgraphs),下面我们来看看由这个方法所预测的自折叠单元与实验的比较。

这里所举的第一个例子为牛胰蛋白酶抑制蛋白。这是一个极小的由 58 个氨基酸残基组成的蛋白质,结构如图 2.2.1(A)所示。在肽链的 N 端和 C 端各有一个螺旋形二级结构,中心部分则由一个长的 β 折叠所组成。图 2.2.1(B)为这个蛋白质结构的表示图及每两个二级结构之间的相互作用(在中心 β 折叠中我们只保留了其中的两个折叠,因为第三个折

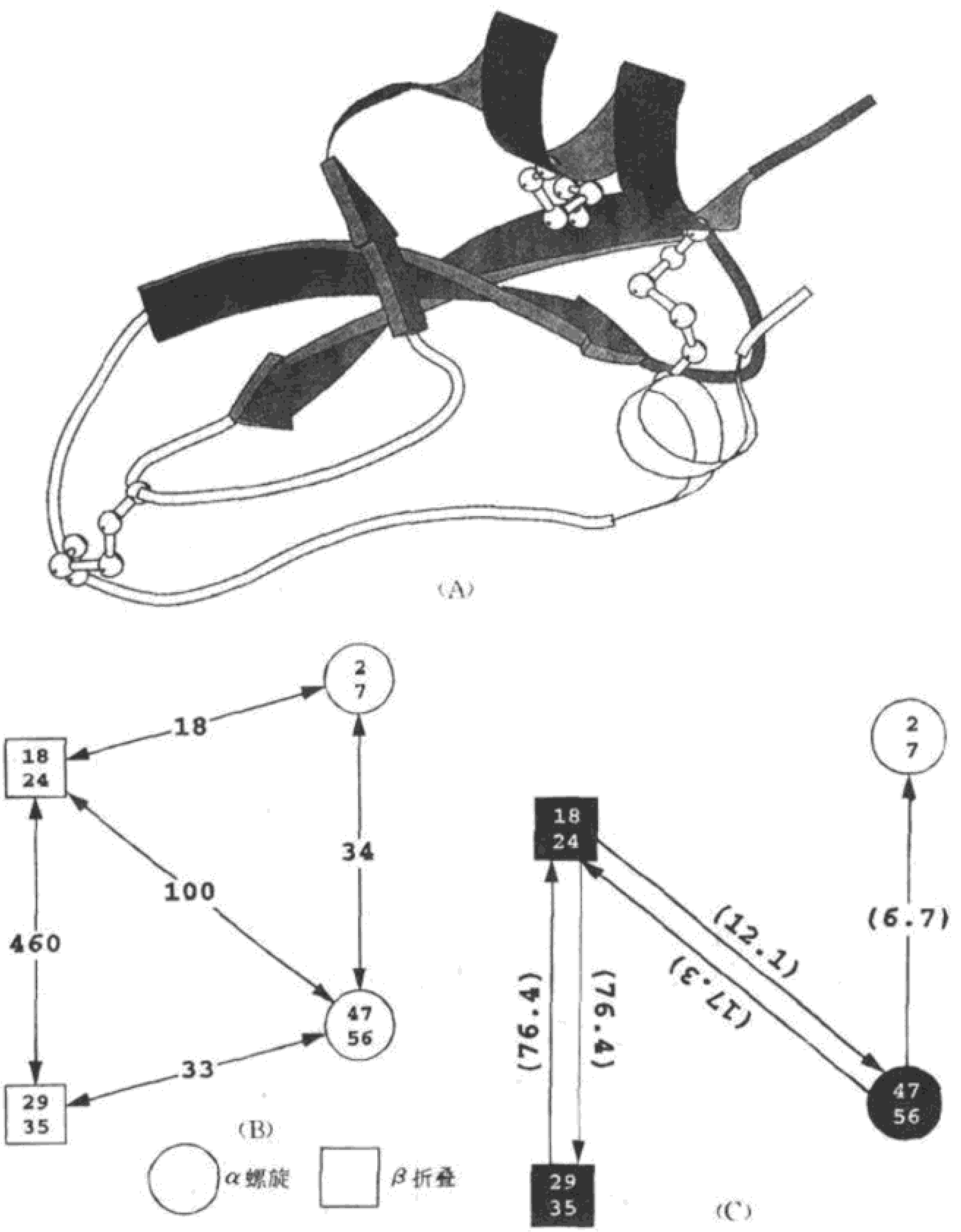


图 2.2.1　牛胰蛋白酶抑制蛋白的晶体结构(PDB5PTI)

(A) 深色部分表示它的一个自折叠单元,又称 $P_\alpha P_\beta-2$。$P_\alpha P_\beta-2$ 由氨基酸序列 18～35 和 44～58 所组成,比早先合成的 $P_\alpha P_\beta$ 有较强的稳定性[3]。

(B) 每一顶点表示一个二级结构,每一条边表示两个二级结构间的相互作用。边上的数字表示两个二级结构间距离在 0.5nm 之间的重原子对的数目。(括号里的数字表示对应的相互作用密度)

(C) 简化图。图中所保留下来的边表示由起点所代表的二级结构对于由终点所代表的二级结构的稳定性起重要作用。注意这个图是非对称的。不难看出,由中心 β 折叠和 C 端 α 螺旋所组成的三个二级结构(深色部分)形成这个简化图的一个最大连通子图。这个连通子图所对应的部分正为 $P_\alpha P_\beta-2$。

叠与前二个之间实际上只有一个氢键相联)。图 2.2.1(C)为简化图,所剩下的边代表这个蛋白质中最重要的相互作用。从图 2.2.1(C)可以看出,N 端的 3_{10}螺旋相对于这个蛋白质的其余部分来说是不重要的。所以中心 β 折叠和 C 端的 α 螺旋应组成一个自治折叠单元,这个预测与实验结果相符合[3]。

第二个例子为嗜热菌蛋白酶(thermolysin)。这是一个较大的蛋白,含有 316 个氨基酸残基。从晶体结构来看,它由 N 端及 C 端两个结构域组成,见图 2.2.2(A)。由 Fontana 领导的研究小组曾对这个蛋白做过长期研究[4]。他们发现 thermolysin 的 N 端结构域不具有独立折叠的能力,而 C 端结构域则可以折叠,除此之外,在 C 端结构域中的某些亚结构域也可以独立折叠。其中最小的一个自折叠单元仅由 61 个氨基酸残基所组成。它的三维结构曾在 1994 年用核磁共振的方法所测定[5]。这个结构与这部分序列在整个蛋白质中所形成的结构有惊人的相似之处。这就证实了这个片断的序列中确实包含了它所形成的结构中所需要的全部信息。

从用我们的方法得到的嗜热菌蛋白酶简化图(图 2.2.2(B)),可以看出,这个蛋白质的 N 端结构域对 C 端结构域有很强的依赖。在两个结构域交界面上的箭头均是由 C 端指向 N 端的,所以我们预测 C 端结构域将能够成为一个自折叠单元,而 N 端结构域则不具有这样的性质。这是与实验结果相符的。不仅如此,从图 2.2.2(B)中还可以看出在 C 端结构域中的一些亚结构域,特别是由残基 259 至 316 组成的最后三个 α 螺旋,也应能够独立地折叠。这正好相当于前面所说的由 61 个氨基酸残基所组成的自折叠单元(256～316)。

总体上,我们目前的方法对于预测已知结构蛋白中的自折叠单元有大约 70%的成功率。在发展这一方法时,我们应用了物理学中的一些思想方法,例如把一个二级结构作为计算的最小单元来考虑(粗粒化)。与以前的一些方法相比较,我们的方法的最大优点是所预测的自折叠单元不需要由连续的氨基酸序列所组成。我们强调理论与实验比较,以及引入尽可能少的能量参数,因为目前对于蛋白质的结构稳定性的了解还很粗浅。一个参数的引进是否必需和它的值的取舍都必须通过与实验比较来加以确认。我们希望通过进一步改进和完善这一方法来对蛋白质的结构的形成原理及折叠的驱动力等一系列基本问题有更深刻的了解。

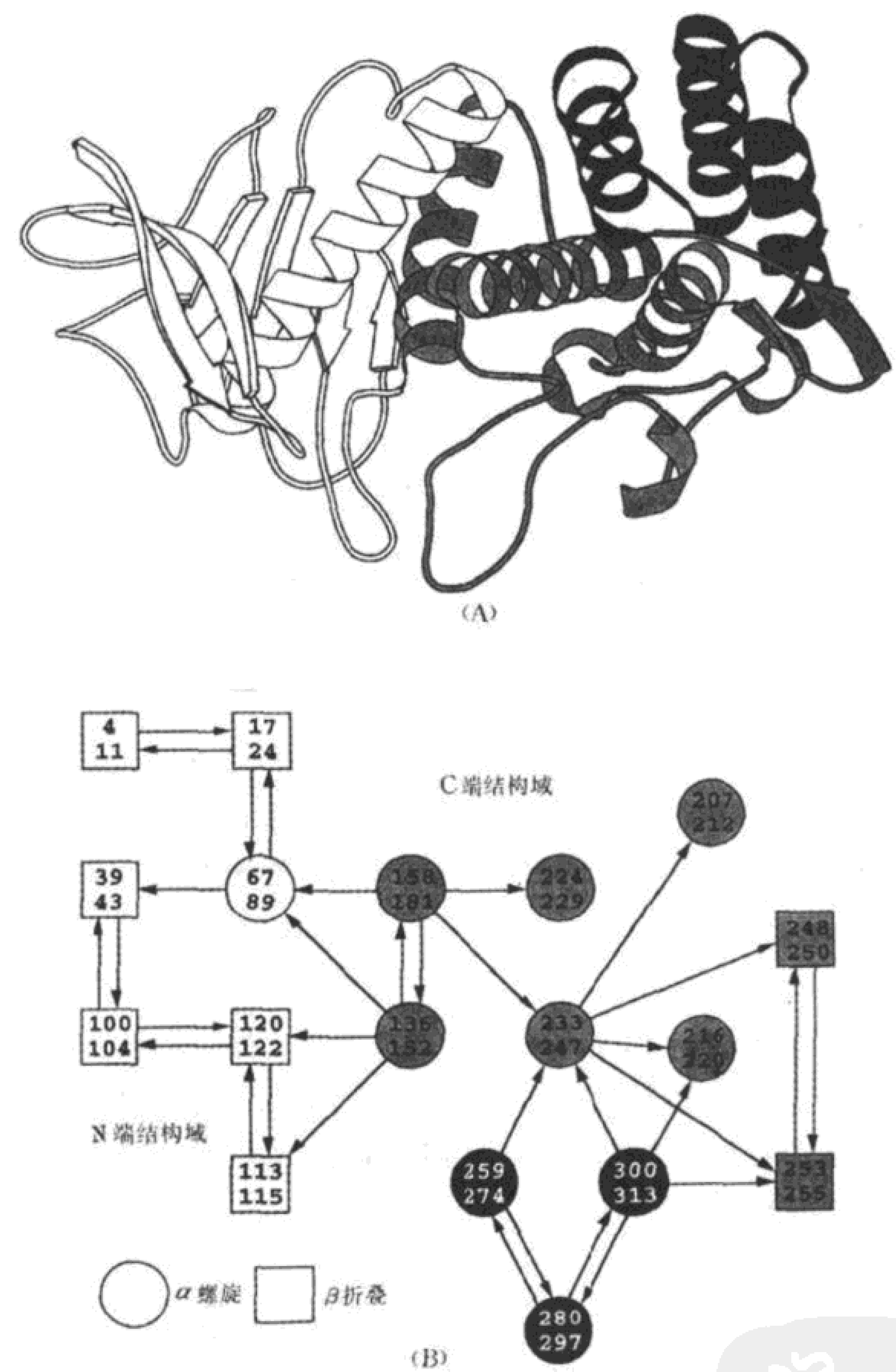


图 2.2.2 嗜热菌蛋白酶

(A) 嗜热菌蛋白酶的晶体结构(PDB3TLN),其中 N 端结构域由残基 1～131 组成,C 端结构域由残基 132～316 组成。Fontana 等发现 thermolysin 的 C 端结构域为一个自折叠单元(灰黑色部分)。不仅如此,由最接近 C 端的 61 个残基所组成的一个亚结构域也形成一个自折叠单元(黑色部分)。这个自折叠单元的结构曾被用核磁共振的方法所测定[4,5]。

(B) 嗜热菌蛋白酶的简化表示图。C 端结构域(灰黑色部分)应为一个自折叠单元,因为在 N 端和 C 端结构域之间的箭头均为由 C 端指向 N 端的。同样,由最接近 C 端的三个 α 螺旋所组成的亚结构域(黑色部分)也应形成一个自折叠单元,因为它们的结构稳定性不依赖于蛋白质的其他部分(没有箭头从蛋白质的其他部分指向这三个二级结构)。这些预测都是与 Fontana 等的实验结果相符合。

参考文献

[1] Srinivasan R, Rose G D. *Proteins: Struct Funct Genet*, 1995, **22**: **81**
[2] Yue K, Dill K A. *Proc Natl Acad Sci* USA, 1995, **92**: 146
[3] Oas T G, Kim P S. *Nature*, 1988, **336**: 42
[4] Fontana A. In: Peptides. Chemistry, Structure and Biology, Rivier J E, Marshall G R, Eds *ESCOM-Leiden*, 1990
[5] Rico M, Jimenez M A, Gonzalez C, *et al*. *Biochemistry*, 1994, **33**: 14834

(彭征宇)

§2.3 分子模拟计算中的理论物理学问题

首先,让我们看一下当前生命科学的基础研究与经济效益甚高的制药工业的密切关系。正像19世纪的机械工业是近代物理学发展的动力一样,制药工业是生命科学发展的主要动力之一,这主要体现在以下三点上。第一,了解自然的目的是为了改造自然。生命科学中的自然通常指动物或植物。药物,无论是由有机分子或生物分子组成,是人们用生物或化学手段来改造自然的主要工具。其次,生命过程的调节及演化是由自然产物包括蛋白质、核酸来完成的。药的组成,则不仅包括所有这些自然产物,而且可以是任何人工合成的分子,因而提供了更为广泛且有效的手段来控制和改造自然的生命过程。最后,随着目前世界范围内的基础科研经费压缩,尽快出产品压力的增加,制药公司提供的合同研究正在成为基础研究工作的一项主要经费来源。

统计资料表明一个主要的新药研制成功通常要花费约15年的时间及5亿美元。这个耗时、耗财的特点促使人们长期以来一直在探索新的方法来优化药物研制过程。粗略地讲,制药过程可由流程图2.3.1表示。图中,最为费时的第一个循环就是寻求与一个给定的生物大分子具有最优结合能的小分子。通常情况下,小分子与大分子结合的部位是知道的。多年研制新药的经验指出,小分子本身的形状总是与大分子上的结合部位的形状互补,如图2.3.2所示。大小分子间的结合能的定义为

$$\Delta G=\Delta E-T\Delta S, \tag{2.3.1}$$

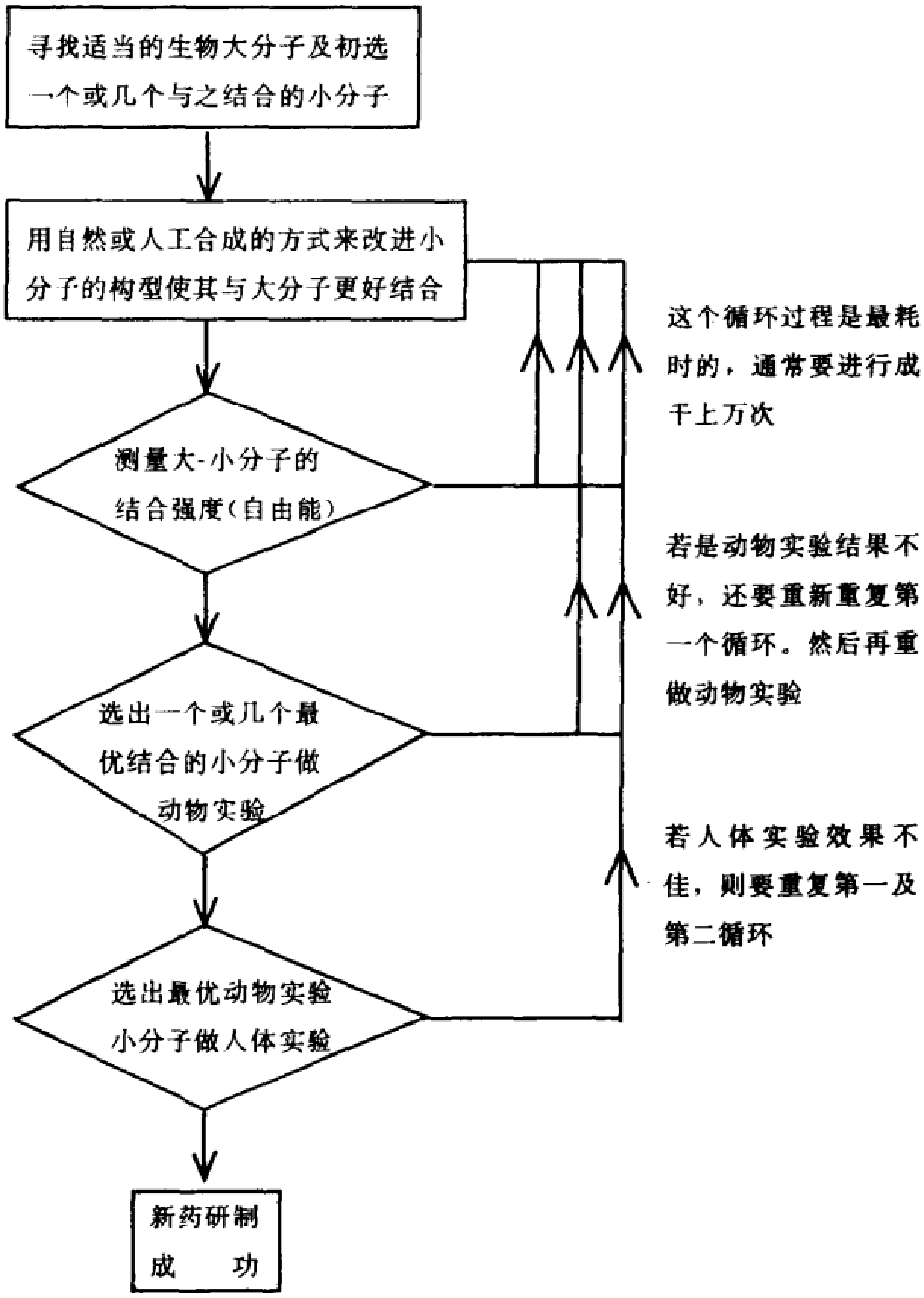


图 2.3.1　制药过程流程图

其中 ΔE 及 ΔS 分别是在大、小分子分开与结合时的能量及熵的差，T 是温度。由此可见，小分子最优化的过程实际上是个物理问题，因为所有涉及的量包括能量、熵、温度、空间形状等都是物理量。因而，若是能从理论上借助于计算机估计出大-小分子的结合能，那将会明显地加快研制新药的过程及缩小研制新药的经费，这也就是所有制药公司都重视计算机辅助分子模拟计算这个理论方法的理由。

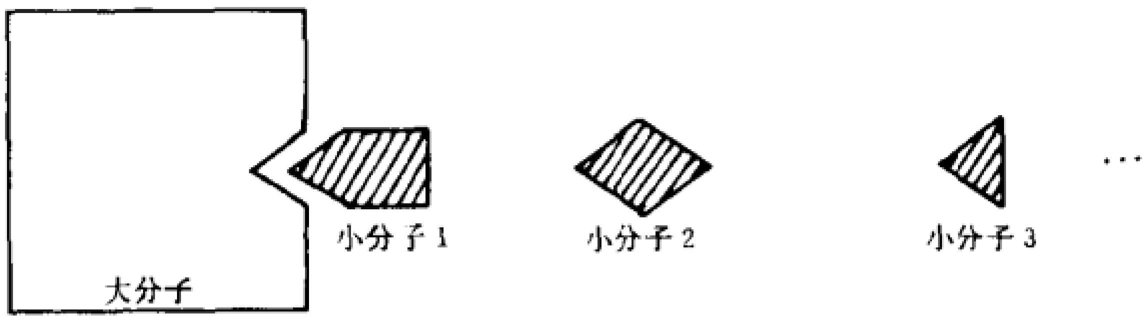


图 2.3.2 小分子与大分子形状互补

应当指出的是，这个大-小分子结合的问题与传统的理论物理问题不同。后者通常研究的要么是只有几个原子的微观系统，要么是亿万个分子的宏观系统。在这个意义上讲，大小分子结合的系统属于"微观大，宏观小"的系统。具体讲，大分子通常是个蛋白质，具有上千个原子，小分子通常有几十到上百个原子。由于大-小分子间的相互作用很敏感地决定于几个或几十个原子间的相互作用，所以通常统计物理中的平均场方法并不提供多少有实际意义的信息。因而必须采用更为细致的方法，需具体到每一个单独原子的描述，这就给理论或计算机计算带来了很大困难。尽管人们通常把这种大-小分子的相互作用作为经典(非量子)的系统来考虑，然而一个具有上千个原子的经典系统的力学问题也不是容易处理的。下面着重分析当前分子模拟计算中的一个主要弊病。通过这个分析可以看出，若是不去发展新的物理学概念及方法，生物计算中的一些主要问题是解决不好的。

在分子计算中的能量函数 E，通常由如下几项组成：

$$E=E_{\text{bond}}+E_{\text{angle}}+E_{\text{dihedral}}+E_{\text{vdw}}+E_{\text{elec}}+\cdots, \qquad (2.3.2)$$

其中 E_{bond} 及 E_{angle} 包括所有由共价键的长度及角度贡献的能量，E_{dihedral} 是由二面角贡献的能量，E_{vdw} 及 E_{elec} 是由范德瓦耳斯力及静电相互作用贡献的能量。除 E_{elec} 外，所有项都属于局部相互作用，因而它们的计算量是与系统中总的原子数成正比的，E_{elec} 的计算量是与原子数的平方成正比的，因而是最费时的。具体地讲，假设蛋白质及小分子分别含有 N 及 n 个原子，其中 $N\gg n\gg 1$。对任何一个给定的蛋白质-小分子构象，E_{elec} 的计算量正比于

$$\frac{N(N-1)}{2}+Nn+\frac{n(n-1)}{2}\sim\frac{N^2}{2}, \qquad (2.3.3)$$

其中$\frac{N(N-1)}{2}$及$\frac{n(n-1)}{2}$分别用于计算蛋白质及小分子的自相互作用项，Nn 用于两者的相互作用项。每一个给定的蛋白质-小分子系统都有很多个构象，总数记作 $m(\gg 1)$，每一个构象的 E_{elec} 计算都要花费$\frac{N^2}{2}$单位。其总的计算量就是

$$m\frac{N^2}{2}+mNn。\tag{2.3.4}$$

仔细看一下就会发现式(2.3.4)说明这种能量计算的方法很不合理，也不经济。因为真正有用的信息是蛋白质与小分子的相互作用及小分子的自相互作用，而计算量却主要花在了计算蛋白质的自相互作用上。遗憾的是这个低效率的、不合理的计算方法，至今仍广泛地应用在分子模拟计算的领域中。

上述这个问题最近有了一个突破。哈佛大学的 Elber 及 Karplus 在 1990 年提出了一个近乎于统计物理中平均场理论的算法，他们称之为 locally enhanced sampling 或 multiple copy simultaneous sampling method，即多构象采样方法。其基本想法是将计算一个真实的蛋白质-小分子系统的问题转化为近似计算一个假想的单蛋白质与多个小分子构象并存的系统。其中每个小分子构象可称为一个“幻影”。这些“幻影”互相“看不见”对方，但每一个“幻影”都与蛋白质正常地相互作用，蛋白质感到的是所有“幻影”的平均影响。在这个意义上讲，这个方法类似于物理中的平均场方法。对于这样的单蛋白质-多“幻影”系统，其 E_{elec} 计算量是正比于下式，

$$\frac{N(N-1)}{2}+mNn+m\frac{n(n-1)}{2}\sim\frac{N^2}{2}+mNn,\tag{2.3.5}$$

其中 m 为 m 个“幻影”数目。式(2.3.5)与式(2.3.4)的最大不同就在于不论计算多少个构象，蛋白质本身的自相互作用只算一次，因而节约了时间。最近，我们把这个算法又推进一步，使之可用来计算一个蛋白质与多个不同构形的不同构象的系统。这个算法的优点有力地显示了理论物理对加快生物计算及研制新药的作用。

最近，我们还认识到多构象采样方法的另外一个优点，尽管我们对它的物理实质还不能定量描述。这个优点是有关熵的信息的。像(2.3.1)中

所述,能量仅仅是结合能中的一项,尽管它是很重要,也许是最重要的一项,还应计及熵项由于熵的计算需要大规模的构象采样,因而费时。通常的计算中只好用很粗略的方式去估算熵,或干脆省去。这也许是很多结合能的计算偏离实验测量结果的一个很主要的原因。我们认为多构象采样本身就已经包含了重要的熵信息。

具体讲,设想初始时,给定小分子的多个"幻影"构象是完全随机的,即处于各种可能的构象中。然后,允许所有构象变化,对这个蛋白质-多"幻影"系统做能量优化,结果是这些"幻影"不断地调整其构象,最终收敛于一个或几个能量最小点,也就是"吸引子"。这些"吸引子"是与构象熵密切联系在一起的。范围越大,熵也越大。因而,做完多构象能量优化后,只要数一下收敛到不同能量极小点的"幻影"数,就可估计熵了。

为了定量地描述,我们引入了收敛强度(degree of clustering):

$$D_{\text{clustering}}=\frac{R_0-R}{R_0}, \tag{2.3.6}$$

其中 R 代表能量优化后,所有构象("幻影")之间的差的均方根值,R_0 代表相应的初始值。有两个极端情况:①若"幻影"们完全不收敛(即系统没有极小能量点),则 $R_0=R_0 \Rightarrow D_{\text{clustering}}=0$;②若系统只有一个极小点,且所有"幻影"收敛到它,则 $R=0, D_{\text{clustering}}=1$。

分母上的 R_0 至关重要。它的引入使收敛性的定义具有内禀性,即它不依赖于一个给定分子本身几何尺寸大小而变。这种内禀性使得我们可以用 $D_{\text{clustering}}$ 来比较不同分子构形(不是构象)的熵变化,因而大大加强了 $D_{\text{clustering}}$ 的实用性。最近,我们用多构象采样的方法及 $D_{\text{clustering}}$ 计算了蛋白质的单点突变问题(它可以看成是一类特殊的蛋白质-小分子系统,因为单点可以被视为一个小分子),并取得了与实验相符合的结果。

以上所举的多构象采样方法的优点尽管都在实际的计算机模拟中得到显示,但它们的理论基础及准确的定量描述仍不十分清楚。首先,多构象采样是一种半平均场近似,到目前为止,对这个近似的可靠性的了解还是很不成熟的。其次,这个采样方法的一个潜在的更大的优点是它能有效地增强能量函数的光滑性。分子模拟用的能量函数与统计物理中的自旋玻璃的能量函数相似,有很多的能量极小点,且能量曲面很不平滑。但不同之处在于分子模拟的能量曲面不平滑性有很大的人为因素。在蛋白质

折叠的研究中有一条最著名的悖论：由于构成蛋白质的肽链具有极多的能量极小点，对应于极多的可能构象，无论怎样都需要近乎无穷长的时间才能从中找出能量最小的构象，也就是自然界中存在的蛋白质的构象。但另一方面，肽链在自然界中只需几秒或更短时间就折叠好了。这说明自然界用的能量函数是相当平滑的。尽管目前已经有了一些计算机摸拟及初步的理论工作证明多构象采样法确实可以有效地改进能量函数的平滑性，其定量描述还远没有建立。我们认为对于这样一个无论是对蛋白质折叠还是计算机辅助制药都极端重要的问题，理论物理学工作者是大有用武之地的，并能给生物学及制药带来实质性的帮助。

参 考 文 献

[1] Elber R, Karplus M. *J Am Chem Soc* USA, 1990, **112**: 9161
[2] Miranker A, Karplus M. *Proteins*, 1991, **11**: 29
[3] Zheng Q, Rosenfeld R, Kyle D J. *J Chem Phys*, 1993, **99**: 8892
[4] Zheng Q, Rosenfeld R, Delisi C, Kyle D J. *Protein Science*, 1994, **3**: 493
[5] Zheng Q, Kyle D J. *Proteins* (accepted)
[6] Zheng Q, Kyle D J. *Bioorganic and Medicinal Chemistry* (accepted)
[7] Roitberg A, Elber R. *J Chem Phys*, 1991, **95**: 9277
[8] Zheng Q, Kyle D J. *Proteins*, 1994, **19**: 324
[9] Zheng W M, Zheng Q. *J Chem Phys*, 1997, **106**: 1191

（郑　强）

§2.4 生物大分子结构的理论分析与模拟

本世纪中叶以来，随着 DNA 双螺旋结构的发现，人们对生命的认识发生了重大飞跃，积累了大量分子水平上有关生命活动规律[1]的知识，这必将对整个生命科学的发展产生巨大的影响。我们自 80 年代初就注意到了这方面的发展，并在生物大分子的序列结构、空间结构和电子结构等方面做了研究。

2.4.1 DNA 序列编码特征的分析

随着世界范围内人类基因组计划(HGP)的开展，很多科学家已经把研究目标转向确定人类基因组中约 30 亿个碱基对(简称 bp，下同)的顺

序;分析人类DNA的序列信息特征;确定5～10万个人类基因的位置。这项研究的深远意义是:由于人类基因组计划将得到存在于DNA中的人体的整套指令,所以人类会对自我有一个根本的了解。它能说明从受精卵如何发育到成熟个体,能说明脑中信息储存与加工的分子基础,能彻底了解巨量的"废物"DNA的功能。这必将对人类健康作出重要贡献,为治疗开辟新的途径。这是因为人们在读出基因组之后,就可了解其长处与短处。很多常见病如心脏病、糖尿病、过敏反应、高血压、精神病和肿瘤都与特定的基因有关,因而可用基因疗法进行预防和治疗。人们还会发现特殊的基因组合与人类的特殊性格、特殊行为的关联,因而有可能创造特定的基因组合以改变人的遗传性状。

用数学、物理学的语言来说,人类基因组计划的最基本、最直接的结果是得到一个由4个元素(T,G,C,A)串接组成的长度为3×10^9 bp的一维链。在这个链上不仅包含有制造人类全部蛋白质的信息(也就是基因),还有按照特定的时空模式把这些蛋白质装配成生物体的四维调控信息(三维空间和一维时间)。如何找到这些信息的编码方式和调节规律是人类基因组研究向数理科学家提出的首要科学问题,而这样巨大的数据如何收集、存储、处理与分配则是人类基因组计划向数理科学家提出的重要技术问题。诺贝尔奖获得者Gilbert在谈到生物学研究模式的变更时指出[2]:"基于全部基因都将知晓并以电子技术可操作的方式驻留在数据库中,新的生物学研究模式的出发点应是理论的。一个科学家将从理论推测出发,然后再返回到实验中去,追踪或验证这些理论假设……生物学家不仅必须成为计算机学者,而且也要改变他们研究生命现象的途径……我们必须把个人计算机进行世界联网,这样才可能了解数据库中每天发生的变化,并适时地彼此进行通讯联系。"正如Gilbert所言,当前人类基因组研究恰恰为我们提出理论假设创造了广阔的天地。

数学家、物理学家在人类基因组研究中的核心作用是解读和利用这本通过测序得到的无标点、无断句,只有4个字符重复30亿次的"天书"。解读后所得到的信息不仅只包含静的结构信息,还应包含动态的四维(时空)调控信息。这一艰巨的工作从何着手呢?最简单的思路是先把这个序列按其功能分割成不同的区域,而要实现这一点最直接的方法是统计。

1. *序列信息的统计分析*[3,4,5]

DNA 上完成不同功能的序列其编码方式不同,因而碱基的使用频率也不同。比如编码蛋白质的区域使用的是三联体码,以相连的三个碱基为一单位。由于蛋白质中 20 种氨基酸出现的频率不同,每个氨基酸三联体码的简并度不同,代表同一氨基酸的几个码的使用频率不同,就造成三联体码的概率分布明显偏离随机分布。如果在 DNA 中寻找到以这种偏好分布着的序列,人们就可以认为它是蛋白质的编码区。国际上有很多工作就是基于这种思想去寻找新基因的。表 2.4.1 给出了我们在人的 22 个基因中统计的氨基酸使用频率的百分数[6]。它充分显示了氨基酸使用频度上的偏好。

表 2.4.1 22 个人类基因的氨基酸使用频率(%)

A	7.63	C	1.20	H	1.82	M	2.30	T	5.83
R	4.89	Q	5.39	I	4.30	F	4.12	W	1.12
N	3.68	E	8.07	L	11.53	P	4.38	Y	2.65
D	5.13	G	5.94	K	6.77	S	6.68	V	6.58

存在于 DNA 序列上的不仅仅是编码区的三联体偏好,目前发现内含子的剪切位点、启动子区、增强子区、各种重复序列等都存在不同编码长度、不同位置上的碱基偏好。我们还发现结构基因的非编码区有同名碱基的类聚现象[7]。这一切都导致了序列不同区域上碱基组成上的差异,从而成为普遍的依赖于"组成"的序列分析方法的基础。

单纯的统计与比较虽然有助于发现未知序列上的典型功能区,但对于阐明 DNA 上信息运作的规律还是远远不够的。为此近年来一些科学家逐渐把若干具有明确物理内涵的方法和参数,特别是非线性科学领域的工具用于 DNA 序列信息分析。我国罗辽复领导的小组[8,9]在这方面作了很多研究。我们小组也使用了分维[10~12]、神经网络[13]、复杂性[14]、局域简并度[15]等指标对结构基因序列的不同区域进行了分析。

2. DNA 序列的分维分析[8,9,10]

我们在分析外显子和内含子序列分维的同时,还引入了碱基间的关联,研究了外显子和内含子序列以双碱基和三碱基为单位的分维。三种情况的研究结果均显示外显子和内含子序列在短程和中程存在自相似性(图 2.4.1),因而可分别定义其分维值 D_g 和 D_m。短程的分维值 D_g 一般比中程分维值 D_m 大,且外显子区的分维值比内含子大。研究结果还表

明，外显子和内含子序列对以2为周期的结构没有位相特异性，而外显子序列对以3为周期的结构显示了位相特异性，即与蛋白质编码一致的第一位相与其他两个位相的分维值不同，但以3为周期的内含子序列并没有显示位相特异性。

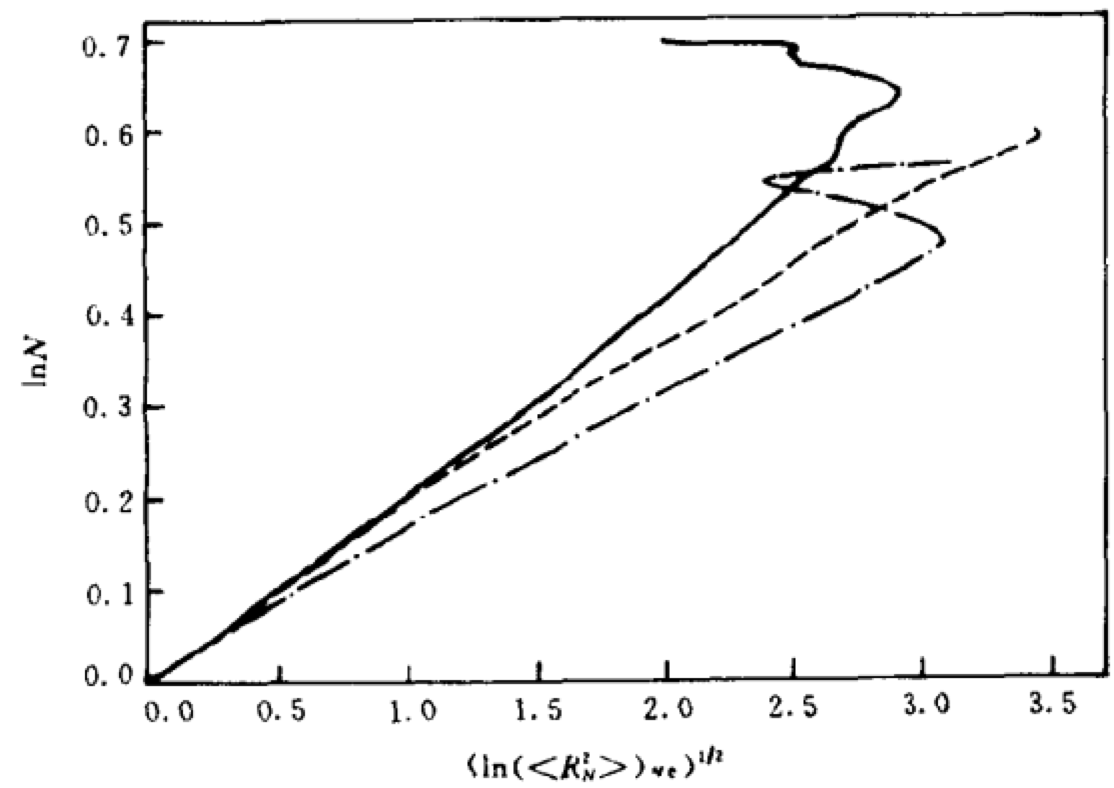


图 2.4.1 外显子和内含子序列的自相似性
实线、虚线、点虚线分别代表DDPSTCAT基因的随机序列、外显子序列和内含子序列。

3. 预测mRNA的剪切位点[13]

我们构造了三层BP网络来预测内含子和外显子间的剪切位点。预测的最大困难是与剪切位点特征序列GT－AG相似的序列在非剪切位点大量存在，它们的数量往往比真正的剪切位点多几十倍。经过仔细选择各种调节因子，神经网络对真实剪切位点的预测成功率可达64％。

4. DNA序列的复杂度分析[14]

我们引用了Lampel和Zip 1976年提出的复杂度概念，测定了DNA序列不同区域的复杂度。我们发现内含子序列、5′侧翼序列以及3′侧翼序列与外显子序列的复杂度值有一定差异，这说明编码蛋白质的DNA序列结构更复杂，倾向于采取更多的配合方式。

5. 突变热点的局域简并度[15]

为了研究氨基酸序列和潜在的功能区之间的关联，我们引入了局域

简并度的概念来分析一个给定的蛋白质片段上每个氨基酸残基的三联体码的简并度。对人凝血因子 IX 的计算发现其突变热点是高简并度区，这显示了两者间的相关。因而此方法可用于预测蛋白质的功能域。

近两三年来 DNA 序列的长程相关[16]是物理学家的一个研究热点，一些研究证明长程相关可以出现在相隔几千个碱基的位点上。

6. 语言学与密码学分析

为了能让系统性的技术或方法介入 DNA 序列分析，科学家们自然想到了语言学和密码学。这是因为既然所有的书都是记载语言所表达的信息的，那么人类基因组序列的这本"天书"也应与某种"语言"相对应，掌握了它的"语言"就可读懂这本书。而密码学在某种意义上可认为是语言在通信传输过程中的安全措施，美国波士顿的科学家们用两个标准的语言学试验分析DNA[17,18,19]，Zipf 试验的结果发现"Junk"DNA 与人类其他语言具有一样的特征，即：单词出现频率的对数与单词排序的对数呈线性关系。Shannon 试验的结果也表明"Junk"DNA 有很大的冗余度，这也和其他人类语言一致。而 DNA 的编码区则不显示上述任何的语言特征。

我们小组曾最早提出用密码学的方法分析 DNA 序列[20]，并用 Friedman 在密码学中使用的重合指数计算了外显子和内含子区域的重合指数值，发现外显子区的数值可比内含子区高好多倍，由此又提出了一种用重合指数预测外显子的方程式。最近国际上的科学家也提到了用密码学方法分析 DNA 序列的有关问题[21]。

当前与 DNA 序列分析相关的基因组信息学正面临这样的态势：①预测的速度正在不断加快，最近有消息表明，整个人类基因组测序工作可望提前 5 年，到2001 年就能完成[22]，因而对"解读"的要求越来越迫切；②基因组相关数据库及电子网络技术高度发达，使得世界各国科学家都能及时得到待分析的资料与数据，也就是说理论研究的基本条件对所有学者是相近的；③占人类基因组 95%左右的"Junk"DNA 的信息运作规律尚未有突破。因而所有从事这一领域研究的科学家都有极好的机会去获得重要进展。我国一些科学家已经注意到了这种趋势，正在努力工作，希望能抓住这一大好时机，做出有份量的工作。为此我们对"Junk"DNA 的功能提出了一些猜想：①可能存在非三联体码，比如长度

大于二的码，像四联码，五联码，九联码 ……等；②这些部分是表达 RNA 的，并通过表达 RNA 构成新的信息表达调控体系；③它的基本功能可能是提供安装生物大分子(如蛋白质等)的四维时空信息，这种信息应有动态特征，并能产生时间标度。

2.4.2 生物大分子的空间结构模拟与药物设计

随着人类基因组计划的执行，找到人类 5～10 万个基因的碱基序列是指日可待的事，因而确定人的上千个原癌基因和几万个与疾病相关的基因表达产物的氨基酸顺序也会逐渐实现。这无疑给人类疾病的治疗带来了极大的希望。然而要了解它们的功能，要找到这些蛋白质致病的分子基础，只有氨基酸顺序的知识是不够的，必须知道它们的三维结构。与此同时，要设计药物治疗这些疾病也需要了解蛋白质的三维结构。这是摆在科学家面前的紧迫任务。当前虽然 X 射线晶体学技术、多维核磁共振(NMR)波谱学技术、二维电子衍射和三维图象重构技术等为蛋白质空间结构测定提供了有效的试验手段，并正以平均每天得到一个生物大分子空间结构的速度前进。但是这些方法提供蛋白质三维结构的速度还远小于蛋白质序列信息的增长速度。另外，这些方法依然存在局限：X 射线晶体学技术需要得到适当的蛋白质晶体；NMR 技术现在还不能有效地测定相对分子质量大于 30 000 的样品；电子晶体学技术目前还正在发展。人们尚不能估计有多少人类蛋白质最终可由上述试验方法测定，又有多少蛋白质不能测定。因此理论模拟与结构预测就显示了重要性，并日益发展和受到人们的重视。理论研究不仅可提供生物大分子空间结构的信息，还能提供电子结构的信息，如能级、表面电荷分布、分子轨道相互作用等以及动力学行为的信息，如生物化学反应中的能量变化、电荷迁移、构象变化等。理论模拟还可研究包括生物分子及其周围环境(如水、离子等)的复杂体系和生物分子的量子效应。上述的不少方面当前是难以直接用分子生物学的实验手段加以研究的。这些模拟的结果对于在分子、亚分子和电子结构层次上了解生命现象的基本过程具有重要意义，为天然生物大分子的改性和基于受体结构的药物分子设计提供依据。

当前分子设计的热点主要集中在两个方面。一方面是对具有明显生物功能的天然蛋白质分子(如胰岛素、天花粉等)加以改性，这是与蛋白质工程紧密联系的。蛋白质工程是本世纪 80 年代初诞生的一个新兴科学领

域。它是以蛋白质的结构规律及其与生物功能的关系为基础，通过分子设计和有控制的基因修饰以及基因合成，对现有蛋白质加以定向改造，为构建并最终生产性能比天然蛋白质更加符合人类社会需要的蛋白质新品种提供科学基础和技术途径，为研究一些重要的分子生物学问题提供强有力的新手段。另一方面是基于生物大分子结构知识的药物设计。传统的药物研制主要是从大量的天然产物，如动物、植物、微生物和合成有机、无机化合物中进行筛选。往往得到一个可供临床使用的药物要筛选1万种不同的化合物，要经过10年左右的时间和耗资2.5～3.0亿美元。当前随着组合化学和群集筛选[23]的发展，这一状况有了一定程度的改善，但耗资、耗时仍然是巨大的。近年来由于结构生物学的发展，相当数量的蛋白质以及一些核酸、多糖的三维结构已被精确了解。到1995年初在PDB (protein data bank，Brookhaven national laboratory)库中收集的生物大分子空间结构数据已达3091套[24]。这就使得基于蛋白质和DNA结构上的药物设计成为可能，为药物开发提供一个新的途径。由此，国际上已发展了一种新的药物设计方法，称为合理药物设计(rational drug design)，这种方法虽不能一次设计成功，但可大大减少实验药物筛选的数量[25]。上述的两个方面既是独立的，又是互相关联的。T. L. Blundell给出的一个图(图2.4.2)[26]可以仔细、全面地说明蛋白质分子设计和蛋白质工程与基于受体的药物设计之间的紧密联系。

分子设计不仅与人类健康关系密切，在其他领域也将发生重要影响。比如在分子电子学和生物计算机领域，正把菌紫质分子作为应用与设计的对象，希望用它组成生物芯片[27,28]。一批工业用酶也正在设计与改性之中，例如：Genetch公司致力于改造枯草杆菌蛋白酶，提高其抗氧化性，以生产含有漂白剂的新一代洗衣粉。丹麦Novo公司用蛋白质工程生产的用于洗涤剂的蛋白酶(dyrazyme)已投放市场。美国的公司正在改造凝乳酶，以解决西方社会广泛食用的乳酪长期保存的问题。用于制造纸浆的木质素酶也正在改造之中。

蛋白质工程和分子设计面临的最大挑战也是最振奋人心的目标，是创建自然界中不存在的优良蛋白质和其他生物活性分子，这方面近年来也已取得了重要进展，被认为已跨越了第一个里程碑[29]。我们小组一直重视生物大分子构象研究，特别关心水溶液环境对生物大分子构象的影

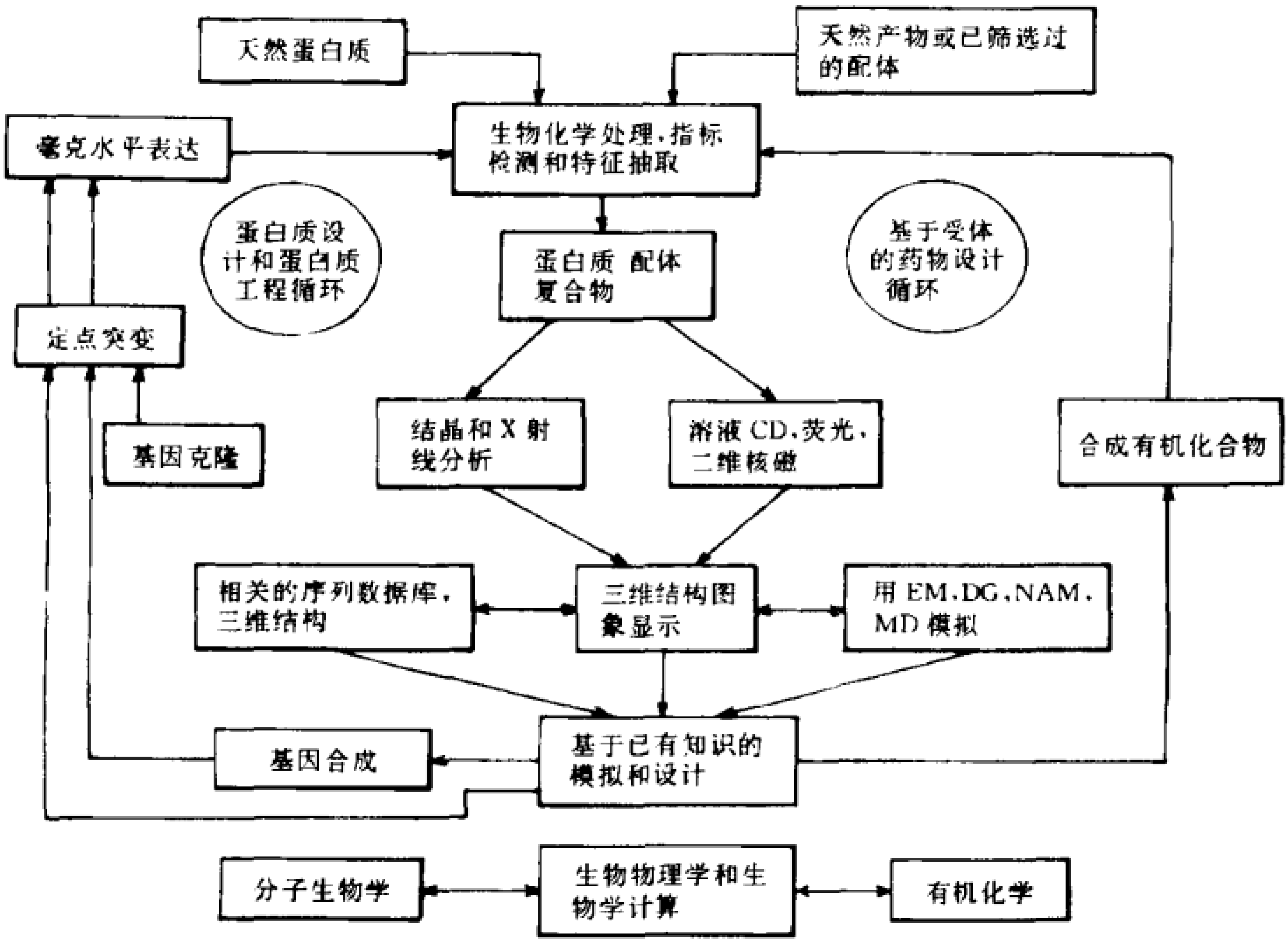


图 2.4.2 蛋白质分子设计示意图

(引自文献[26])

响。为了实现从理论上得到生物大分子溶液构象的目的,我们选择了由小分子到大分子,由无水环境到有水环境这样的由简单到复杂逐渐过渡的工作步骤。为了适应各步工作的需要,我们先后建立了如下方法:

① 水和氨基酸相互作用势的计算(包括从头计算及非线性拟合);

② 生物分子水合等能图分析;

③ Monte Carlo 模拟;

④ 分子动力学(包括随机动力学)模拟;

⑤ 自由能微扰计算;

⑥ 神经网络的结构预测。

1. 小分子体系中结构模拟的研究

在小分子体系中我们主要进行了如下几项工作:

(1) 在从头计算的精度上拟合了水和丙氨酸,缬氨酸两性离子的相互作用势[30~34]。这类势的建立将用于模拟生物大分子水溶液结构。

(2) 以小肽 antamanide 和脑啡肽为对象研究了它们的水合等能图[35,36,37],找到了亲水区、疏水区和水合位点。

(3) 用 Monte Carlo 模拟方法研究了丙氨酸[38]、缬氨酸[39]和环二肽[L-His-L-Thr][40]体系的水溶液结构和能量特征，以此为大分子水溶液结构研究作了尝试。这些工作在国际上也是开展得较早的。

(4) 建立了适用于自由能微扰计算的部分力偏歧的 Monte Carlo 方法，并用于由甲烷到己醇变化体系的自由能计算[41,42]。这一方法为研究定位突变后生物大分子能量、构象的变化打下了基础。

2. 生物大分子的构象模拟及结构、功能关系的研究

几年来我们对有明显生物功能，有大量实验资料的几个典型大分子体系进行了理论研究，所得的结果是很有意义的。

(1) 胰岛素研究[43,44,45]

为了配合我国在胰岛素晶体结构分析上取得的优异成果，我们以胰岛素为对象作了如下三个工作：

1) 计算了胰岛素分子的水合等能图，得到了胰岛素各部分的亲水、疏水特征，并与晶体结构分析所给出的水分子位置进行了比较，发现两者有很大的一致性。

2) 对胰岛素 B 链 4 号位的 Gln 的两种可能构象进行了理论判别，给出了其中一种构象是优势构象的结论，为 X 衍射方法所不能确定的构象补充了信息。

3) 用分子动力学模拟方法，经过长达 600 ps 的模拟研究了去七肽胰岛素的晶体结构，给出了它的理论模拟构象，这一结果将与 X 射线晶体分析结果进行比较。

(2) 血红蛋白的构象模拟[46]

虽然生物大分子构象的理论研究工作不少，但在模拟退火方法出现前使用 Monte Carlo 方法的工作却几乎没有。原因是在局部极小的搜寻中遍历性不好，因而很难得到整体极小。为了克服这一困难，我们使用了反复热浴的方法，研究并比较了正常与镰刀形红细胞贫血病血红蛋白片段的构象。发现正常血红蛋白片段的模拟构象与晶体衍射结果一致，而镰刀形红细胞贫血病血红蛋白片段的构象并没有因 Glu-Val 交换而发生明显变化，从而支持了镰刀形红细胞贫血病的病因是由于电荷改变，进而引起血红蛋白分子聚集这一假说。

(3) 细胞色素 C 的随机动力学模拟[47]

我们对细胞色素 C 活性中心附近一段肽链的运动借助 Langevin 方

程等进行了计算。结果表明,尽管有证据说明在还原态与氧化态之间该肽段的构象变化是细胞色素C完成其生物功能所要求的,但它不是柔性造成的随机运动的结果。

(4) 胰凝乳蛋白酶激活过程的分子动力学研究[48]

对这一系统我们进行了长达纳秒数量级的多次模拟,发现从初态到终态,中间需要克服一个位垒,并确定了这一位垒的位置。同时也证明了这一激活过程的时间不可能少于纳秒数量级。这些结果对进一步研究酶激活过程是十分有意义的。

(5) 菌紫质研究(bacterio rhodopsin,简称BR)[49,50,51]

由于近年来BR与生物芯片的关系密切,而引起了人们的高度重视。我们主要做了如下两方面的工作:

1) 对BR构象进行了分子动力学的模拟,并在对模拟结构仔细分析的基础上,指出了原初光物理过程推动了视黄醛顺反异构化,从而导致质子泵出现的推论,因而提出了视黄醛的顺反异构化模型。

2) 建立了蛋白质二级结构预测的神经网络的方法,并用此方法预测了典型的膜蛋白BR和视紫质(rhodopsin)的二级结构,所得的结果与实验资料一致,从而为膜蛋白的结构预测提供了可能。

3. *神经网络方法在结构预测中的应用*

自1989年开始,我们就构建了BP网络,并不断加以改进。先后将此方法用于各种构象模拟工作中去。

(1) 预测蛋白质二级结构[51]

我们根据Qian和Sejnowski的文章[52]建立了预测蛋白质二级结构的神经网络,并对此方法的误差以及造成误差的可能原因进行了深入的分析,在此基础上提出了分布矩阵的概念和一种修正的学习方法,使得α螺旋及β折叠预测的精确度和相关系数均有所提高。

(2) 预测同源蛋白的三级结构[53]

我们构建了预测同源蛋白C_α原子距离限制的神经网络,并将这些限制用于分子动力学模拟,以得到蛋白质的三级结构。通过选取一组输氧蛋白作为研究对象,证实了方法的实用性,其距离限制的预测成功率可达90%。

(3) 预测tRNA的二级结构[54]

我们建立了神经网络与碱基配对规则相结合的方法, 用于预测

tRNA的二级结构，得到了满意的结果，其预测成功率在88%以上，优于其他预测方法。此方法有可能进一步用于rRNA和mRNA的二级结构分析中。

我们还将神经网络方法用于膜蛋白二级结构预测[50]和蛋白质主链二面角的分析[55]，也得到了一些有意义的结果。上述工作说明神经网络方法是一种结构模拟研究的有效辅助方法。

4. 其他与生物大分子构象相关的研究

(1) 旋转势垒的计算[56]

本工作是从量子力学的角度探讨核苷发生构象变化的物理本质。通过对与Z-DNA相对应的鸟苷的旋转势垒的计算得到其沿糖苷键的旋转势垒为6.69 kJ/mol，因而在这个位置发生构象变化是可能的。

(2) 生物分子的手性相互作用[57,58]

我们曾经计算了L-L和L-D丙氨酸二体的相互作用能。在给定条件下L-L体系的能谷比L-D体系更深，两者能量差为50.66 kJ/mol，显示L-L二体有更大的稳定性。有关电荷分布，键级和偶极矩的计算结果也说明L-L二体有更强的键合趋势。

(3) DNA的解旋模型[59]

基于对核苷酸构象及旋转势垒的分析，我们曾提出一个DNA局部构象左旋化的解链模型，并探讨了该模型在信息表达，调节中的作用。

随着结构生物学的发展和药物设计的需要，人们对生物大分子结构模拟的要求是迫切的，这一领域的发展也是迅速的。当前影响其进一步发展的最突出的困难可以归结为两方面：

(1) 物理模型过于粗糙：在大多数分子设计中都采用经验势函数，这种把一个多粒子体系量子力学水平上的相互作用，以原子对之间相互作用的和来表示的方法显然是过于简单了。用这样的力场虽然可以处理包含几千到一万个原子的体系，但模拟结果在多大程度上是真实的，始终是令人怀疑的。

(2) 如何找到整体极小：通常认为生物分子的稳定构象是自由能最低的构象，因此模拟的目的就是要找到与最低自由能相对应的构象。但是在一个含有成百上千个原子的体系中可能存在的局域极小点是个天文数字，迄今为止尚没有足够有效的办法来达到这一点[31]。

当前科学家对克服这些困难是充满信心的。

力场改进的主要制约因素是计算能力，因为严格地说从本世纪30年代发展起来的量子物理理论对描述多粒子体系是严格的，其后50年代发展的Hartree-Fock-Roothaan方法，密度泛函方法等，在处理分子体系时虽有一些近似，但物理模型还是合理的。只是囿于生物分子粒子数过多，计算能力不允许才不得不使用经验势函数。当前速度达每秒几十亿次浮点运算的机器已商品化，百亿次、万亿次的机器正在试验，随着计算机能力的快速提高，使用复杂力场的可能性越来越大，物理模型会越来越真实。近几年发展的量子力学和分子力学(或分子动力学)结合的方法，就是这种趋势的体现。当然在这个过程中还有很多技术问题要解决。

极值搜寻问题在目前仍旧是一个没有解决的问题，数学家也没有给出确定性的办法。但是对它的讨论与改进是很多的[31,32]。要根本解决这一问题，了解真实地发生在生物体中的事件也许是重要的。蛋白质尽管自由度极大，但在生物体中最终都形成了它的确定的天然态，可以说多极值问题在活体中是解决了的。所以要确定蛋白质的构象应该研究蛋白质的折叠过程，要把生物学的信息加到理论模拟中去。现在蛋白质分子设计中有效的同源构筑方法，就是依赖于结构生物学提供的信息。近年来对蛋白质构象模式的研究取得了很大进展[33,60,61]。比较普遍的看法是：蛋白质的折叠类型是有限的，目前估计为几百至几千种。这要远小于蛋白质所具有的自由度数目。同时蛋白质的折叠类型是与它们的组分和一级序列相关的，这样就有可能从蛋白质的初级信息中确定它们的最终折叠类型。如果能把这些来自生物学的信息加到理论模拟系统中去，也许是解决这一问题的好办法。也正是这一点令很多科学家感到鼓舞并充满希望。

展望今后，从本世纪末到下世纪初的20年间，蛋白质及药物设计将处于一个迅速发展的时期，一方面研究方法和技术将日臻成熟，另一方面可望产生一批性能优良的蛋白质工程制剂和药物，在医疗卫生、工业、农业等方面应用。在此基础上分子设计和蛋白质工程作为新兴高技术领域，向产业化过渡的过程也将开始，这将对未来世纪的社会生产和生活带来重大影响。

2.4.3 生物大分子电子结构研究

从量子力学可以清楚地了解电荷与空间矢量总是联系在一起的，因

此在用量子力学方法定量分析生物大分子的时候电子结构和空间构象总是相辅相成一起出现的。因此我们在探讨生物大分子空间结构的同时，一直坚持研究这些分子的电子结构。在生物分子电子结构研究中，除了计算一些典型分子体系的电荷分布、键级、偶极矩等参量外，我们集中分析了蛋白质与核酸的能带，并在此基础上进一步研究了水和无机离子对能带结构的影响。接着完善了 Green 函数方法，把局部杂质对能带的影响这一复杂情况又加了进去，以使得对生物大分子电子结构的理论研究逐渐接近真实系统。为了比较理论与实验结果，我们又在两个比较简单的体系上计算了与电子结构直接相关的理论俄歇谱。

1. 生物大分子体系能带结构的研究[62~73]

在国内用半经验方法研究的基础上，1985 年以来，我们与德国 Nuernberg-Erlangen 大学的 J. Ladik 教授等科学家一起将 Monte Carlo 方法、互洽场(mutually consistant field)方法和从头计算的晶体轨道法(abinitio CO)结合起来，发展了一种精确地研究环境(水、离子等)对生物大分子电子结构影响的新方法，并用此方法在国际上首次全面系统地研究了水和无机离子对α，β 构象的蛋白质以及DNA 能带的影响。

能带研究所得的主要结果是：

(1) 水和离子的存在改变了各体系导带与价带的位置；

(2) 水和离子的存在改变了各体系的禁带宽度(band gap)，但改变的方向因氨基酸和核苷酸的种类而异。禁带宽度的最大减小量约为 2 eV；

(3) 离子的存在能强烈地引起水-水相互作用能及水-氨基酸相互作用能的下降，这表明离子的存在削弱了水与水以及水与氨基酸的相互作用关系，从而改变了水溶液的结构。

2. Hartree-Fock Green 矩阵法研究[74,75]

在研究了环境(水和离子)对生物大分子电子结构的影响之后，进一步的深入就是其他小分子物质(如致癌剂等)在局部对生物大分子的影响。在技术上这是一个十分复杂的问题。近年来国际上正发展一种称作 Hatree-Fock Green 矩阵法的新方法，它是把现有的从头计算分子轨道法(abinitio Mo)，从头计算晶体轨道法(abinitio CO)和 Green 函数法结合起来，用于处理高聚物的局域微扰问题。这一方法不仅在生物学领域有广阔的发展前景，将来可用于致癌剂引起的 DNA 点突变和蛋白质工程中定位突变这样一类十分重要的问题。而且在物理学和化学等其他领域也

有重要和明显的实际意义。我们在 Li－H 准一维系统得到了收敛的结果，给出了结合态，共振态的特征以及收敛极限。

3. 纯水及水溶液体系的俄歇谱研究[76,77]

俄歇电子谱是当前测量技术中的一项常用方法，俄歇电子跃迁是物质电子结构的反映，因而可用量子物理学方法加以研究。我们与 C. M. Liegener 博士等把 Green 函数方法和 Monte Carlo 方法结合起来，在国际上首次从理论上得到了溶液体系的俄歇谱。从我们所用的两个体系(甘氨酸两性离子水溶液和纯水体系)看来，结果是非常好的。甘氨酸两性离子水溶液的理论俄歇谱与相应的实验资料能很好地吻合。纯水体系目前虽还没有实验资料，但我们将其与水蒸气和冰的实验结果进行了比较，发现液态水的理论谱更接近冰的实验谱，这与预期的结果是完全一致的。

4. 鸟苷的电子结构计算[78]

我们计算了与 Z－DNA 和 B－DNA 相应的鸟苷的碱基原子净电荷，并研究了净电荷在碱基旋转过程中的变化，发现随着 DNA 种类不同或随着碱基的旋转，原子净电荷也发生不同性质的变化，其中以 N7 的净电荷变化最大，这可能影响重要的生物功能。

生物体是多层次(分子、细胞、组织、器官、系统)的复杂系统，为了阐明生命活动的本质，我们在把主要力量集中于研究生物大分子序列和结构的同时，也注意使用非线性动力学的耗散理论分析生物大分子体系的特征[79~82]。在这一领域我们提出了蛋白质合成的动力学模型，并研究了这一模型的演化及分子构型对其稳态动力学结构的影响。这一工作的目的是希望从分子水平来探讨生物体形态发生的本质。

近年来国际上生物学与物理学、数学的结合越来越紧密，越来越本质。彭桓武教授早就注意到这一发展趋势，十几年来我们小组在彭先生的指导、鼓励与支持下做了一些工作，今后我们要更好地坚持这一方向，以实际的研究成果表示对彭先生深切的谢意与敬意。

参考文献

[1] 陈润生. 物理学与社会. 北京：北京大学出版社，1992：142

[2] Gilbert W. *Nature*, 1991, **349**: 99

[3] Gribskov M, Devereux J. Sequence Analysis Primer. New York: Stockton Press, 1991

[4] Doolittle R F. Methods in Enzymology. San Diego: Academic Press,1990

[5] Bishop M J, Rawlings C J. Nucleic acid and protein sequence analysis. Oxford: IRL Press, 1987

[6] 徐军,凌伦奖,沈如群等. 生物物理学报,1992,**8**: 711

[7] 凌伦奖,徐军,沈如群等. 生物物理学报,1994,**10**: 174

[8] Lou L F, Tsai Li, Zhou Y M. *J Theor Biol*, 1988, **130**: 351

[9] 周宇明,罗辽复. 生物物理学报,1992,**8**: 581

[10] Xiao Yi, Chen Runsheng, Shen Ruqun, Sun Jian, Xu Jun. *Protein Sequence & Data Analysis*, 1993: 415

[11] Xiao Yi, Chen Runsheng, Shen Ruqun, Sun Jian, Xu Jun. *J Theor Biol*, 1995, **175**: 23

[12] 肖毅,陈润生,沈如群等. 生物物理学报,1993,**9**: 641

[13] 孙键,徐军,凌伦奖等. 生物物理学报,1993,**9**: 127

[14] 沈如群,陈润生,凌伦奖等. 科学通报,1993,**38**: 1995

[15] Xu Jun, Chen Runshen, Xiao Zhi-xiong. *Analytical Biochemisty*, 1994,**23**: 71

[16] Maddox J. *Nature*, 1993,**358**: 103

[17] Flam F. *Science*, 1994,**266**: 1320

[18] Mantegna R N, *et al.* *Physical Review Letters*, 1994, **73**: 3169

[19] Brendel V, *et al.* *Nucleic Acids Research*, 1984, **12**: 2561

[20] Xu Jun, Chen Runshen, Ling Lunjiang, Shen Ruqun, Sun Jian. *Comput Biol Med*, 1993, **23**: 333

[21] Henikoff S. In: Biocomputing (D W Smith Ed). San Diego: Academic Press, 1994. 87

[22] Marshall E. *Science*, 1995, **267**: 783

[23] Bevan P, Ryder H, Shaw I. *TIBTECH*, 1995, **13**: 115

[24] Protein Data Bank. *Quarterly Newsletter*, *Jan* 1995 (Release #71)

[25] 来鲁华等编. 蛋白质的结构预测与分子设计. 北京:北京大学出版社,1993

[26] Blundell T L. *TIBTECH*, 1994, **12**: 145

[27] Birge R R. *Scientific American*, 1995, **272**: 90

[28] 细菌视紫红质的原型研究与生物芯片,中科院生物物理所理论室译,中国生物物理学会,1988

[29] Ho S P, DeGrade W F. *J Am Chem Soc*, 1987, **109**: 6751

[30] 凌伦奖,倪向善,石秀凡,王化军,陈润生. 生物物理学报,1990,**6**: 202

[31] Wolynes P G, Onuchic J N, Thirumalai D. *Science*, 1995,**267**: 1619

[32] Cvijovic D, Klinowski J. *Science*, 1995, **267**: 664

[33] Blundell T L, Johnson M S. *Protein Sci*, 1993, **2**: 877

[34] Ni Xiangshan, Shi Xiufan, Ling Lunjiang. *Int J Quantum Chem*,1988, **34**: 527

[35] Chen Run-sheng, Ni Xiang-shan, Shi Xiu-fan. Water and Ions in Biological Systems. New York: Plenum Press, 1985

[36] Shi Xiu fan, Chen Runsheng, Ni Xiangshan. *J Mol Sci* (Wuhan China), 1985, **3**: 243

[37] 岳琳,石秀凡,倪向善等. 生物化学和生物物理学报,1985,**3**: 334

[38] 倪向善,陈润生,石秀凡. 中国科学(B辑),1985, **1**: 45

[39] 倪向善,陈润生,石秀凡. 生物物理学报,1985,**4**: 227

[40] 岳琳,石秀凡,倪向善等. 生物物理学报,1986,**1**:64
[41] 石秀凡,倪向善,凌伦奖. 生物物理学报,1989,**5**:246
[42] 倪向善,石秀凡,陈润生. 生物物理学报,1993,**9**:309
[43] 石秀凡,凌伦奖,倪向善等. 科学通报,1986,**31**:376
[44] 倪向善,陈润生,石秀凡. 科学通报,1986,**31**:936
[45] 王化军,陈润生,倪向善等. 生物物理学报,1986,**4**:390
[46] Chen R S. *J Comput Chem*,1989,**10**:488
[47] 张学军,陈润生,倪向善. 生物物理学报,1985,**1**:43
[48] 王化军,凌伦奖,石秀凡等. 生物物理学报,1990,**6**:217
[49] 凌伦奖,石秀凡,倪向善等. 生物物理学报,1989,**5**:293
[50] 陈润生,王化军,石秀凡等. 生物物理学报,1990,**6**:267
[51] 王化军,陈润生,倪向善等. 生物物理学报,1989,**5**:422
[52] Qian N, Sejnowski T J. *J Mol Biol*, 1988, **202**: 865
[53] 孙键,凌伦奖,陈润生. 高技术通讯,1991, **4**:1
[54] 孙键,凌伦奖,陈润生. 高技术通讯,1991,**8**:1
[55] 王化军,石秀凡,凌伦奖,陈润生. 生物物理学报,1991,**7**:157
[56] 陈润生. 分子科学学报,1982,**3**:81
[57] 陈润生,倪向善. 分子科学学报,1982,**4**:191
[58] 陈润生,倪向善,彭旱元. 生物化学与生物物理学进展,1982,**5**:35
[59] 陈润生. 分子科学学报,1982,**1**:81
[60] Orengo C A, Jones D T, Thornton J M. *Nature*, 1994,**372**:631
[61] Chou K C, Zhang C T. *J Biol Chem*, 1994, **269**:22014
[62] Xiao Yi, Chen Run-sheng. *Int J Quantum Chem*, 1985, **28**:573
[63] Chen R S, Otto P, Ladik J. *J Chem Phys*, 1986, **85**:5365
[64] Liegener C M, Otto P, Chen R S, Ladik J. *Theoretica Chimica ACTA*, 1988, **73**: 449
[65] Chen R S, Liegener C-M, Otto P, Ladik J. *ACTA Biochim Biophys Hung*, 1987, **22**: 205
[66] 肖奕,陈润生. 科学通报,1985,**18**:1418
[67] 肖奕,陈润生,石秀凡. 生物物理学报,1985,**1**:47
[68] 肖奕,陈润生,倪向善. 生物物理学报,1986,**1**:51
[69] 肖奕,陈润生. 生物化学与生物物理学报,1986,**18**:138
[70] 陈润生,石秀凡,倪向善,凌伦奖. 生物物理学报,1988,**4**:370
[71] 陈润生,倪向善,石秀凡,凌伦奖. 生物物理学报,1989,**5**:35
[72] 肖奕,陈润生. 生物化学及生物物理学进展,1986,**1**:54
[73] Liegener C M, Chen R S, Otto P, Ladik J. *Coll Czeck Chem Comm*, 1988, **53**: 1946
[74] Chen R S, Otto P, Seel M, Ladik J. *Chem Phys*, 1988, **120**:177
[75] 陈润生,李晓春,倪向善. 科学通报,1990,**8**:621
[76] Liegener C M, Bakhshi A K, Chen R S, Ladik J. *J Chem Phys*, 1987, **86**:6039
[77] Liegener C M, Chen R S. *J Chem Phys*, 1988, **88**:2618
[78] 陈润生. 分子科学学报,1982,**4**:195

[79] 陈润生. 中国科学(B辑),1982,**7**:629
[80] 陈润生,倪向善. 科学通报,1982,**8**:500
[81] 倪向善,陈润生. 科学通报,1982,**7**:430
[82] 陈润生,倪向善. 分子科学学报,1982,**1**:123

(陈润生)

§2.5 生物分子的计算机模拟和结构、动力学、热力学

自1953年沃森和克里克提出DNA双螺旋模型以来,生物学的发展已进入了一个全新的阶段,即分子生物学的时代。现代生物学不再是一门描述性的学科,而成为一门像物理学和化学那样精确的定量科学。分子生物学是在分子水平上阐明生命现象规律的学科,它的诞生将物理学、化学与生命科学有机地联系在一起。正如A. L. Lehninger在他著名的生化教科书中所说:“生物体与所有自然过程遵从相同的物理规律。”物质世界是统一的,但另一方面生物大分子又有其自身的特点:首先分子量大,含有成百上千个原子;结构复杂,没有对称性;原子与原子间相互作用不可忽略;生物活性与特定的空间结构相关,这特定的空间结构主要靠弱键(范德华相互作用、氢键、盐键)所维系;分子周围的溶剂十分重要;分子有柔性和运动性等。生物大分子的特殊性,要求人们对其物理学规律的认识进一步加深和发展。

长期以来对生物大分子的研究主要用的是实验方法,这包括生物化学方法、分子遗传学方法、免疫学方法等,以及各种物理手段,如X射线晶体衍射,多维核磁共振波谱,电子自旋共振,红外光谱、紫外光谱、荧光光谱、激光拉曼光谱,穆斯堡尔谱以及扫描隧道电子显微镜、原子力显微镜、中子衍射、电子衍射等。自70年代中期以来由于计算机的发展,计算机模拟已成为研究生物分子结构、动力学与功能关系的另一个重要手段。

2.5.1 计算机模拟

所谓计算机模拟是一种计算机实验,它是从原子水平上的相互作用出发,借助计算机数值模拟的方法得到生物大分子结构、动力学、热力学方面的信息,以及研究这些信息与生物功能之间的关系,如图2.5.1所示。

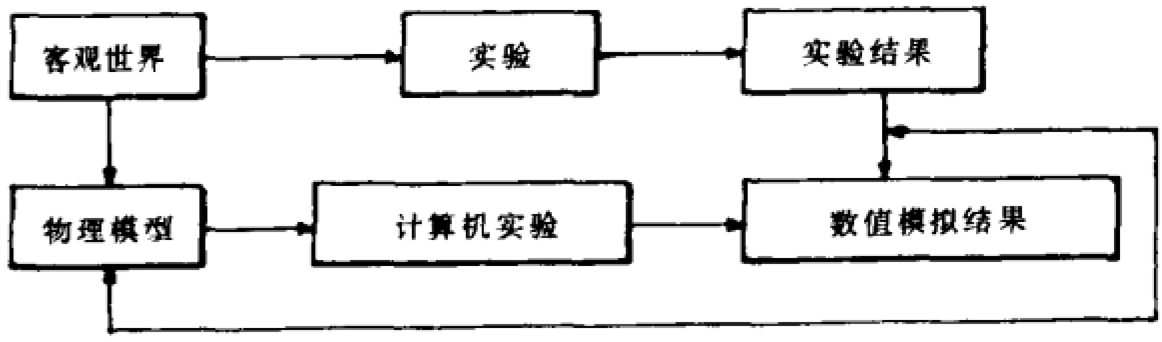


图 2.5.1 生物大分子的计算机模拟示意图

通过计算机模拟可以探测实验难以探测到的性质，可以与实验结合解释实验结果，以及预测结构变化引起的新的性质。

具体的研究内容可以包括：酶与底物及抑制剂的相互识别、相互作用；酶的反应机理、反应速率；蛋白质折叠及蛋白质的稳定性；蛋白质及其周围溶剂的静电相互作用、疏水相互作用，蛋白质的水化；光合反应中心及其他氧化还原体系中长程的电子转移；离子通道中的离子输运。

《自然，结构生物学》(*Nature, Structural Biology*)(1994 年 1 月)前言中有这样一段话："随着结构测定越来越容易，人们强调的重点将从描述转移到功能分析，尽管在任何科学努力中描述是重要的最初阶段。什么力驱使蛋白质与其底物结合？在离体和活体情况下蛋白质如何折叠成天然状态？是什么因子决定了蛋白质的稳定性？蛋白质如何将化学能转变成力？……"结构生物学所研究的不仅仅是结构，更多关心的是生物分子的功能。

2.5.2 分子动力学模拟和随机动力学模拟

1. 原理

计算机模拟的方法很多，用于大分子的最重要的方法是分子动力学模拟(MD)和随机动力学模拟(SD)[1]。1977 年哈佛大学 M. Karplus 首先将此方法用于蛋白质。分子动力学模拟的基本原理是：在准各态历经假设下，任何力学量对系综的平均，等于该力学量对时间的平均，而力学量对时间的平均可从经典运动方程所决定的运动轨迹得到。具体地说，模拟从已知的或假设的蛋白质的空间坐标出发，给每个原子赋以一定的初速度，解牛顿运动方程：

$$m_i \frac{d^2 r_i(t)}{dt^2} = F_i(t),$$

$$F_1=\frac{\partial V_i(r_1,r_2,\cdots,r_N)}{\partial t},$$

其中 $V_i(r_1,r_2,\cdots,r_N)$ 为半经验的势函数。通过数值求解牛顿运动方程得到一个时间序列 $\{r_i(t_1),r_i(t_2),\cdots,r_i(t_n)\}$，即可得到任一物理量的系综平均：

$$\langle Q\rangle=\overline{Q}_{(t)}=\frac{1}{t}\int_0^t Q(r(t'))\mathrm{d}t'。$$

由于用 MD 方法模拟蛋白质在水溶液中时，大量计算往往用于计算水-水相互作用。因此可将溶剂对蛋白质的作用，用溶剂对蛋白质的平均作用势，随机碰撞力和摩擦力来近似，即用朗之万方程代替牛顿方程，

$$m_i\frac{\mathrm{d}^2r_i(t)}{\mathrm{d}t^2}=F_i(t)-m_i\int_0^t\zeta(t-\tau)\frac{\mathrm{d}r_i(\tau)}{\mathrm{d}\tau}\mathrm{d}\tau+R_i(t),$$

或用布朗运动方程代替牛顿方程，

$$m_i\frac{\mathrm{d}^2r_i(t)}{\mathrm{d}t^2}=F_i(t)-m_i\gamma_i\frac{\mathrm{d}r_i(t)}{\mathrm{d}t}+R_i(t)。$$

这就是随机动力学模拟。通过分子动力学模拟[2]或随机动力学模拟[3]可以得到蛋白质体系的平均能量：总能量、动能、势能；平均结构：原子平均位置 $\langle r_i\rangle$，平均二面角 $\langle\varphi_i\rangle$；动力学信息：原子位置的均方根涨落 $\langle(\Delta r_i)^2\rangle^{\frac{1}{2}}$，二面角的均方根涨落 $\langle(\Delta\varphi_i)^2\rangle^{\frac{1}{2}}$；原子位置涨落的自相关函数 $C(\tau)=\langle\Delta r_i(t)\Delta r_i(t+\tau)\rangle$ 等。若在哈密顿量 $H(P,r)$ 中引入微扰参量，再通过热力学积分或热力学微扰的方法还可以得到热力学信息：自由能和熵[4]。

分子动力学模拟和随机动力学模拟已成为研究蛋白质、核酸结构与功能的有力工具[5,6]。

2. 应用举例

自由能和熵是与生化实验关系最密切的两个物理量。但是这两个物理量不能够直接由计算机模拟得到，其原因如下。

发现一个体系处于微观状态 (r,P) 的概率是：

$$\pi(P,r)=\frac{\exp[-H(P,r)/K_BT]}{h^{3N}N!\ Z}。$$

任何一个物理量 Q 的系综平均：

$$\langle Q\rangle=\iint Q(P,r)\pi(P,r)\mathrm{d}P\mathrm{d}r,$$

对于 N 个原子的体系，经典哈密顿 $H(P,r)=\sum_{i=1}^{N}\frac{P_i^2}{2m_i}+V(r_1,r_2,\cdots,r_N)$，其中坐标 $r=(r_1,r_2,\cdots,r_N)$，动量 $P=(P_1,P_2,\cdots,P_N)$，正则系综的亥姆霍茨自由能和熵分别为：

$$F(N,V,T)=-K_BT\ln Z(N,V,T)=-K_BT\ln\langle\exp(\beta H)\rangle,$$

$$\begin{aligned}S(N,V,T)&=\frac{\langle H\rangle}{T}-K_BT\ln Z(N,V,T)\\&=\frac{\langle H\rangle}{T}-K_BT\ln\langle\exp(\beta H)\rangle,\end{aligned}$$

其中 $\beta=\frac{1}{K_BT}$。

因为 $\langle\exp(\beta H)\rangle=1+\beta\langle H\rangle+\cdots$，由公式 $\ln(1+x)=x-\frac{x^2}{2}+\cdots$ 可知 $\ln\langle\exp(\beta H)\rangle=\beta\langle H\rangle-\frac{\beta^2}{2}\langle H\rangle^2+\cdots$。因而位形空间的高能区对自由能和熵有大的贡献，而寻常分子动力学模拟或蒙特卡洛法主要在位形空间的低能区采样，为此必须在哈密顿量中引入耦联参量 λ，

$$H(P,r,\lambda)=\sum_{i=1}^{N}\frac{P_i^2}{2m_i(\lambda)}+V(r_1,r_2,\cdots,r_N,\lambda),$$

$$m_1(\lambda)=(1-\lambda)m_i^A+\lambda m_i^B,$$

$\lambda=0$，为 A 态；$\lambda=1$ 为 B 态。

通过热力学微扰法

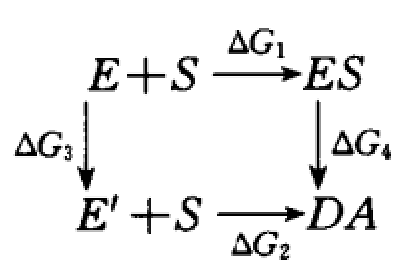
$$\Delta F_{BA}=-K_BT\ln\langle\exp[-\beta\Delta H_{BA}(P,r,\lambda)]\rangle_A,$$

或热力学积分法

$$\Delta F_{BA}=\int_{\lambda_A}^{\lambda_B}\langle\frac{\partial H(P,r,\lambda)}{\partial\lambda}\rangle_\lambda d\lambda,$$

由计算机模拟求得。

在蛋白质分子设计及药物分子设计中，人们感兴趣的是比较野生型的酶 E 和突变体的酶 E' 与底物结合自由能的差，或者比较两种不同药物 D 和 D' 与受体 R 结合自由能之差。因此可以通过如下的热力学循环：



$$\begin{array}{ccc} D+A & \xrightarrow{\Delta G_1} & DA \\ \Delta G_3 \downarrow & & \downarrow \Delta G_4 \\ D'+A & \xrightarrow[\Delta G_2]{} & D'A \end{array}$$

计算 $$\Delta\Delta G=\Delta G_1-\Delta G_2=\Delta G_3-\Delta G_4。$$

我们用上述方法比较了枯草杆菌蛋白酶野生型和突变体(G169A)与底物结合自由能之差[7]，柔红霉素与其类似物及 DNA 片断结合自由能之差[8]，以及葡萄糖异构酶野生型与突变体(N184D)与底物结合自由能之差[9]，并从结构的角度对结合自由能之改变进行了解释，另外还分析了自由能模拟中的困难及存在的问题[10]。此篇文章，得到了国际上的关注(编者评论，*Protein: Structure, Function and Genetics*, 1995, **17**(1))。

为了使自由能计算有更广泛的应用，刘海燕最近成功地发展了用单一参考态来模拟自由能之差的方法，该方法可能使自由能计算用于多个药物的选择以及多个突变体的选择。

2.5.3 量子效应

尽管经典的分子动力学模拟已取得很大成功，但是当问题涉及到酶反应，包括化学键的断裂和生成以及涉及到光合反应中心或其他氧化还原体系中的长程电子转移时，就必须考虑量子效应，经典的分子动力学模拟不能解决这个问题。而现有的量子化学计算方法也不可能处理生物大分子这么大的体系。另外，按照化学反应速率理论，速率常数与反应通量时间相关函数有关，所以人们不仅对静态的性质感兴趣，而且希望能作动力学模拟。因此发展能考虑量子效应的分子动力学模拟方法已成为当前人们关心的热点。解决这个问题的方法之一就是用量子力学来处理少数自由度，而大部分自由度用经典力学来处理，同时还需考虑两部分自由度之间的耦合。

木糖异构酶是工业上用来生产高果糖浆的一种工业用酶，它的作用是催化葡萄糖异构化为果糖，或木糖异构化为木酮糖。目前该酶与底物复合物的高分辨率的晶体结构已经得到。该酶由 2～4 个亚基组成，每个亚基中有 2 个二价金属离子(Mg^{2+}、Mn^{2+}或 Co^{2+})，一个起稳定结构的作用，另一个起催化作用。其中起催化作用的金属离子又有 A,B 两个结合

位点，二价金属离子是酶的活性所必需的。目前认为酶的催化机理是负氢离子转移机理。但是金属离子在此催化机理中究竟起了什么作用并不清楚。

我们将半经验的量化程序包 MOPAC 与经典分子动力学模拟程序包 GROMOS 结合形成了一个可作动力学模拟的程序包[11]。

系统哈密顿量：　$H = H_{QM} + H_{MM} + H_{QM/MM}$。

其中酶活性中心部分：$2Mg^{2+}$，一个 H_2O，Glu 181，Glu 217 的 γ - coo^-，Asp 245，Asp 255，Asp 257，Asp 287 的 β - coo^- 以及 Lys 183 - CH_2 - NH_3^+，His 220 的咪唑环及底物 D - Xylose 用半经验量化 PM_3 计算，酶的其他部分用经典力学 GROMOS 势。

假设的异构化反应按以下五个反应步骤进行：

① 与 Mg^{2+} 配位的水分子将一个质子转移给 257 位的 Asp 的羧基，形成 OH^-；

② OH^- 进攻木糖 C^2 的 OH 基团生成 H_2O 和氧负离子；

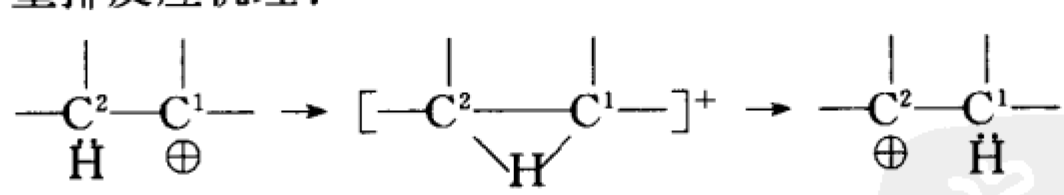
③ 木糖 C^2 上的氢原子以负氢离子的形式转移到 C^1 上，C^2 与 O^2 形成羰基，C^1 由 sp^2 杂化变为 sp^3 杂化，O^1 形成氧负离子；

④ Lys 183 将 NH_3^+ 上一个质子给了 O^1 负离子使之形成羟基；

⑤ Lys 183 获得一个质子。

我们分别计算了起催化作用的 Mg^{2+} 在 A 位和 B 位时沿反应路径的能量剖面。并与真空中用 PM3 的计算结果进行了对比。我们对每一点先作 400 步能量极小化，再作 5 000 步分子动力学模拟，然后再作能量极小化至能量改变小于 0.0001 kJ/mol。

按照有机化学中关于负氢离子转移的 Meermein - Pondorff - Verley - Openaue 重排反应机理：



C^1 上的正电荷越多越有利于 C^2 上的负氢离子转移。

计算结果表明，Mg^{2+} 起始在 A 位，反应过程中从 A 位移动到 B 位，它作为一个 Lewis 酸，使 C^1 - O^1 键极化，使 C^1 上的电荷由 0.42 变为 0.51，O^1 上的负电荷由 −0.56 变为 −0.72，从而有利于 C^2 上的氢作为负

氢离子从 C^2 转移到 C^1，Mg^{2+} 起了催化作用。另外在反应过程中 Lys183 的 $\varepsilon-NH_3^+$ 和 D-Xylose 的 C^1-O^1 相互接近，而 H_2O 和 D-Xylose O^{-1} 的距离当 Mg^{2+} 在 A 位和 B 位时分别为 0.369 nm 和 0.27 nm。我们认为 Lys183 直接参与了反应，是质子给体，而不像前人所假定的 H_2O 是质子给体，Lys183 只起了稳定结构的作用。我们的结论与基因定点突变的结果一致。在酶环境下的计算机模拟结果与真空中的计算结果对比将可以说明酶的作用。

2.5.4 蛋白质及其周围溶剂的静电相互作用

蛋白质与其周围溶剂的静电相互作用十分重要，目前国际上处理静电问题的方法尽管很多，不外乎是宏观模型和微观模型两种[12]。宏观模型又称连续介质模型，将蛋白质内部和周围溶剂视为连续介质。内部用泊松方程，外部用泊松-玻尔兹曼方程描述。

$$\begin{cases} \nabla^2\Phi_1(r) = -\dfrac{1}{\varepsilon_1}\sum\limits_{i=1}^{N} q_i\delta(r-r_i), & \text{泊松方程，} \\ \nabla^2\Phi_2(r) = \dfrac{2eIN_2}{\varepsilon_2}\sinh\left(\dfrac{e\Phi_2(r)}{K_B T}\right), & \text{泊松-玻尔兹曼方程。} \end{cases}$$

在低离子强度下，非线性的泊松-玻尔兹曼方程可近似为线性泊松-玻尔兹曼方程：

$$\nabla^2\Phi_2(r) = K^2\Phi_2(r), \qquad K^2 = \frac{2IF^2}{\varepsilon_2 K_B T},$$

在一定边界条件下，求解

$$\text{边界条件：}\quad \begin{cases} \Phi_1(r) = \Phi_2(r), \\ \varepsilon_1\nabla_n\Phi_1(r) = \varepsilon_2\nabla_n\Phi_2(r)。 \end{cases}$$

由于蛋白质的不规则形状，一般都需数值求解，使用有限差分法、有限元法或边界元法等方法。

连续介质模型的优点是计算量小，参数少。所遇到的困难是如何确定蛋白质内部的介电常数。

微观模型已广泛用于分子力学和分子动力学计算，其中最著名的微观模型是 Warshel 提出的 PDLD 模型，该模型考虑了蛋白质的部分电荷、诱导偶极以及溶剂分子的固有偶极矩。微观模型的缺点是计算量大。

上述两种模型均已成功用于研究蛋白质水化，计算蛋白质活性中

心可解离基团的 pKa 值，蛋白质的氧化还原电势，蛋白质的稳定性等。

我们研究了用边界元方法解蛋白质及其周围溶剂静电场问题的方法，并尝试了将边界元法与 PDLD 模型结合[13]以及将 PDLD 模型与 Tanford Kirk Wood 理论结合计算 BPTI 四个残基的水化自由能和枯草杆菌蛋白酶 BPN′突变体可解离基团 His64 的 pKa 值。还研究了用线性方法解非线性泊松-玻尔兹曼方程的方法[14]，以及边界元方法中的关键问题：蛋白质表面三角面元的划分方法[15]。

参考文献

[1] *Computer Simulation of Biomolecular Systems, Theory and experimental application*, 1989, **1**, 1993, **2**, *ESCOM Leiden*

[2] Shi Y Y, Yun R H, Van Gunsteren W F. Molecular dynamics simulation of despentapeptide Insulin in Crystal enviroment. *J Mol Biol*, 1988, **200**: 571

[3] Shi Y Y, Wang L, Van Gunsteren W F. On the approximation of solvent effects on the conformation and dynamics of cyclosporin A by stochastic dynamics simulation techniques. *Molecular Simulation*, 1988, **1**: 369

[4] Beveridge D L, Dicapua F M. *Ann Rev Biophys Biophys Chem*, 1989, **18**: 431

[5] McCammon J A, Harvey S C. Dynamics of proteins and Nucleic Acids. London, Cambridge University press, 1987

[6] Brooks III C L, Karplus M, Pettitt B M. *Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics, Adv Chem Phys* Vol 61. John Wiley and Sons, NY, 1988

[7] Wang C X, Shi Y Y, Zhou F, Wang L. Thermodynamic integration Calculation of binding free energy difference for Gly 169 mutation in subtilisin BPN′. *Proteins, Structure, Function and genetics*, 1993, **15**: 5

[8] Shi Y Y, Zhao H M, Wang C X. Relative binding free energy calculation of DNA to Daunomycin and it's 13 - dihydro analogue. *International Journal of Biological Macromolecules*, 1993, **15**: 247

[9] Hu H, Shi Y Y, Wang C X. Exploring the interaction between D - xylose isomerase and D - xylose by free energy calculation, *Protein*. 1996, **26**: 157

[10] Shi Y Y, Mark A, Wang C X, Huang F H, Berenson R J H, Van Gunsteren W F. Can the stability of protein mutants be predicted by free energy calculations. *Protein engineering*, 1993, **6**(3): 289

[11] Liu H Y, Shi Y Y. Combination of Molecular Dynamics Simulation program GROMOS with Semi - empirical Quantum Chemistry Method AMI. *Chinese Bulletin*, 1993, **38**(23): 2023

[12] Smith P E, Van Gunsteren W F. Methods for the evaluation for long range electrostatic forces, in Computer Simulations of Molecular Systems, *Theory and experimental application ESCOM Leiden*, 1993, **2**: 182

[13] Xiang Z X, Huang F H, Shi Y Y. Calculation of Solvation Energy with a Combination of Boundary Element Method and PDLD Model. *J Phy Chem*, 1994, **98**(48): 12782

[14] Xiang Zhexin, Xu Yingwu, Shi Yunyu. Solving the Finite - Difference Non - Linear Problem under a Linear Approach. *J Comp Chem*, 1995, **16**(2): 200

[15] Xiang Z X, Xu Y W, Shi Y Y. Calculating the Electric Potential of Macromolecules a Simple Method for Molecular Surface Triangulation. *J Comp Chem*, 1995, **16**(4): 512

(施蕴渝)

§2.6 复杂生物体系的理论计算和药物设计

现代生物学的发展,已经从生物圈、生态、群落、个体、系统、器官、组织和细胞的研究层次(传统生物学),深入到分子、亚分子(分子生物学和结构生物学)和电子结构的层次(量子生物学)[1]。生命现象的分子基础和微观机制的研究现已成为一个非常活跃和具有挑战性的研究领域。这一领域的研究,采用了生物、物理、数学、化学和计算机科学等多学科交叉融合的手段,以及实验研究和理论研究相结合的方法。其中,对复杂生物大分子体系及其相互作用进行理论计算和计算机模拟,占有特别重要的地位,它对于在分子、亚分子和电子结构的层次上了解许多生命现象的本质与规律,具有重要的意义,在某些方面已成为实验难以替代的研究手段[2,3]。

许多重大的生命现象,如遗传、生长、发育、进化以及病理和药理作用,均与生物体内分子间的相互作用有关[4]。生物体内分子间的相互作用包括生物大分子和生物大分子间的相互作用以及生物大分子(受体)和活性小分子(配体)间的相互作用,其中配体小分子又包括内源性的神经递质、信使分子、激素和外源性的化合物(如药物)。

配体参与、影响、调节神经、心血管和免疫等各个系统的生理活动,配体与受体生物大分子的相互作用研究,很自然地成为现代科学研究的前沿之一。随着分子生物学的发展,大量的活性肽和功能蛋白质被分离、纯化和鉴定,许多受体基因也已被克隆和表达,有的已测定了其三维结构。用分子生物学突变技术以及物理、化学方法阐明了

许多配体和受体生物大分子的分子识别和结合位点，积累了丰富的有关配体和受体分子间相互作用的实验数据，亟需有深入的理论研究与之配合。

配体与受体的相互识别和配体对受体的激活作用，是配体产生特异性生物作用的基础。配体在体内作用的受体，通常是一些生物大分子(蛋白质、酶、核酸等)，它们的三维结构，在很大程度上决定了对于一个有效的配体分子结构的要求。因此，从生物大分子的三维结构知识出发，研究配体与受体的相互作用，从分子、亚分子和电子结构的水平探索配体-受体作用和构效关系的规律，具有重要意义。例如：①研究受体生物大分子上配体结合部位的结构和性质；②配体-受体复合物的构型和立体化学特征；③配体与受体结合的模式和选择性(特异性)；④配体分子与受体结合时的活性基团和活性构象等等。这将有助于深入揭示生命现象的本质和分子机制，同时也为具有特定性能生物活性分子的设计提供理论依据，因此不仅具有学术价值，也具有一定的应用价值[5]。

随着保护知识产权的有关法规在我国逐步实行，新药的创制已日益显示出其重要性和紧迫性。然而迄今为止，寻找新药仍然是一件耗费大而效率低的工作，一个重要的原因就是缺乏深入的理论指导。据统计，目前世界上要筛选和试验 10 000 种以上的化合物，需花费 3～5 亿美元的投资，才能发现一种成功的药物。这种状况迫切需要采用新的理论方法和技术加以改进。药物在体内的作用过程，与受体生物大分子(酶、膜蛋白、离子通道等)的结构与功能密切相关。因此，应用各种理论计算方法和分子模拟技术，进行药物分子或其他生物活性分子的设计研究，对于推动我国的新药创制工作也具有十分重要的意义。

由于生物体系的极端复杂性，在相当长的时期内，生物大分子体系的理论计算与分子模拟一直是一个十分困难的问题。然而，近年来，随着许多新的理论方法和计算技术的发展以及计算机的飞速进步，这种状况已有了很大改变：

① 已经可以采用量子化学从头计算方法计算近 400 个原子的 DNA 片段[6]；

② 可以采用分子力学和分子动力学的方法来模拟核酸、蛋白质和多糖等生物大分子的结构以及它们之间的相互作用[7,8]；

③ 最近又发展了量子力学和分子力学、分子动力学相结合的方法来研究生物大分子体系的相互作用、酶催化机理以及生物体系的电子、质子、离子和能量传递等生命科学中的重要问题[3,9,10]。

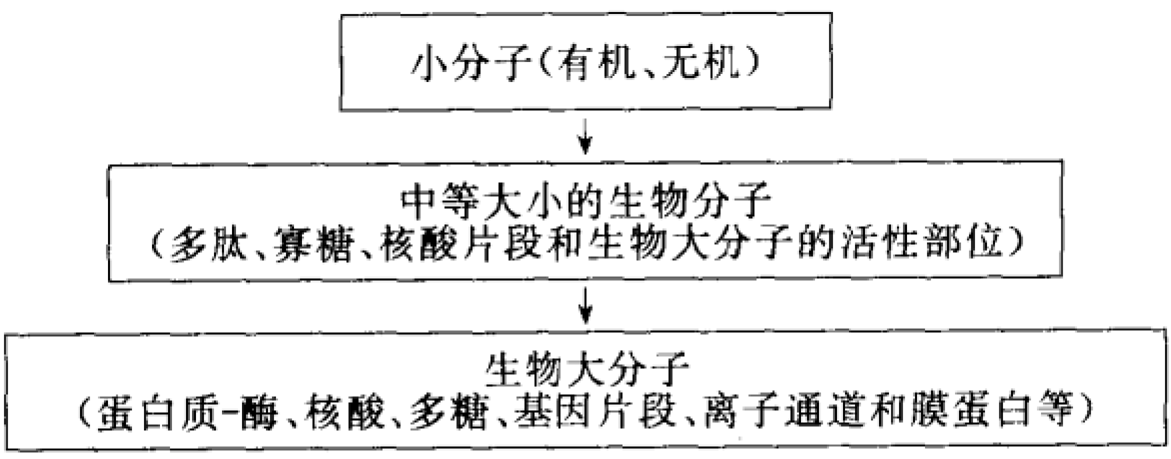
运用理论计算方法来研究复杂的生物大分子体系，现已成为国际上十分活跃、发展很快的研究领域。第七、第八届国际量子化学会议都专门设立分会来研讨生物大分子的理论计算和基因工程及蛋白质工程中的分子模拟问题。同时，在理论计算与分子模拟的基础上，生物活性分子(包括药物分子)的设计，在实际应用中也有了很大突破。

本文就近年来国际上关于复杂生物体系的理论计算方法和分子模拟技术及其在药物设计中的应用作一简单的介绍。

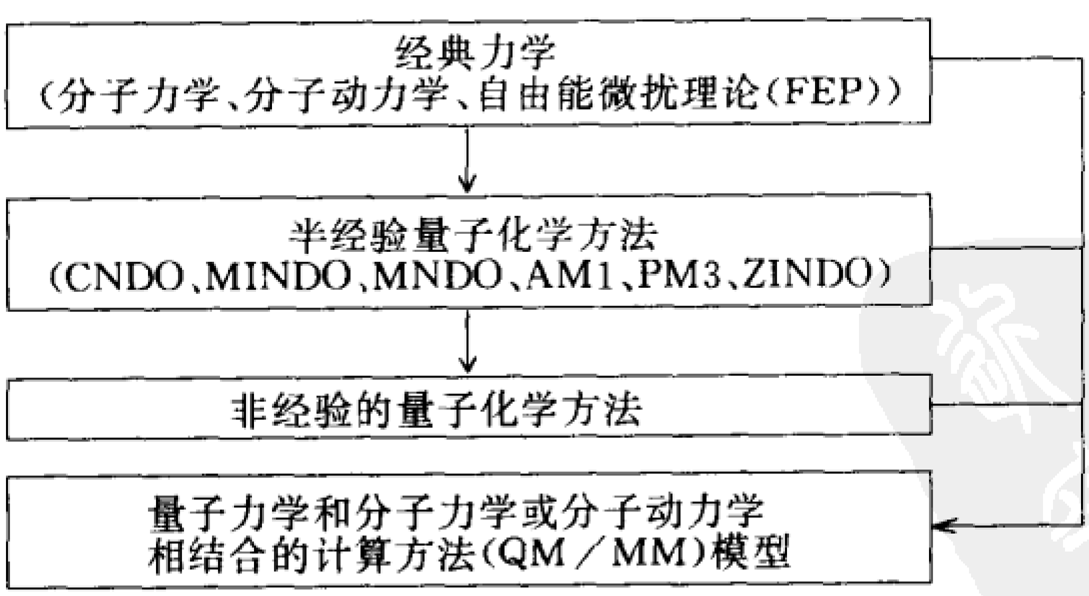
2.6.1 理论计算方法

随着计算机和计算方法的不断发展，理论化学计算方法及其应用有了很大的发展，并且在生命科学中的应用也越来越广泛。总的来说，生物大分子体系的理论计算有如下几大趋势：

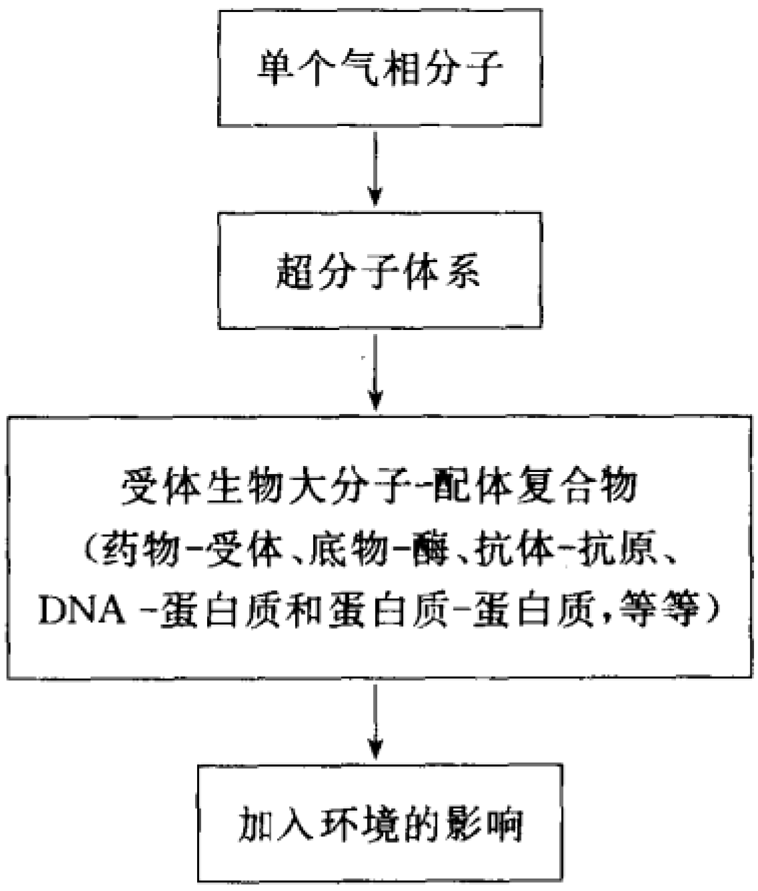
(1) 所计算的分子越来越大：



(2) 所用的方法越来越精确：



(3) 所计算的体系越来越复杂：



下面就生物大分子体系中的理论计算方法及其应用作一简单的介绍。

1．分子力学方法(MM)

分子力学应用经验势函数，即力场方法模拟分子的结构，计算分子的性质。通常计算生物大分子的经验势函数为[9]：

$$\begin{aligned} E_{\text{total}} = & \sum_{\text{bonds}} K_r (r - r_{eq})^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_{eq})^2 \\ & + \sum_{\text{dihedrals}} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] \\ & + \sum_{i<j}^{\text{nonbonded}} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\varepsilon R_{ij}} \right] \\ & + \sum_{H-\text{bonds}} \left[\frac{C_{ij}}{R_{ij}^{12}} - \frac{D_{ij}}{R_{ij}^{10}} \right], \end{aligned} \tag{2.6.1}$$

其中的 K_r、K_θ、V_n、A_{ij}、B_{ij}、C_{ij}和 D_{ij}为力场参数。力场参数可通过理论计算或实验数据拟合得到，根据经验势函数形式及拟合体系和方法的不同，可得到不同的力场参数，即不同的力场，表 2.6.1 列出了现常用于有机分子和生物大分子计算的力场。

表 2.6.1 有机分子和生物大分子体系计算常用的力场

力场	应用范围	文献
MM2/MM3	有机小分子	11,12
Amber	生物大分子	13,14
CHARMM	生物大分子	15,16
Tripos	有机小分子和生物大分子	17,18
YETI	生物体系的分子,特别是含有金属的生物大分子	19～21
Universal	所有类型的小分子(主族元素化合物,有机分子和金属配合物)。对元素周期表中的所有元素都进行了参数化	22～25
Dreiding	主族元素组成的有机小分子和无机小分子	26
GROMOS	生物大分子	27,28
POLARIS	生物大分子	29

应用分子力学可进行生物分子的结构优化、构象分析、配体-受体的相互作用计算等研究工作。

2. 分子动力学方法(MD)

MD 的理论基础是,分子在构象变化时,其热力学平衡构象所占的机率最多。MD 方法是在给定分子势函数和力场的情况下,用解经典力学牛顿方程的方法研究分子的运动和构型空间,其基本假设是各态历经假说:无限时间平均等于系综平均或对整个构型空间的积分。MD 可以计算生物大分子的结构和热力学性质,可以模拟生物大分子在溶液中或其他环境中的构象[30]。

在 MD 计算时假设分子中原子的运动符合牛顿定律,则分子的运动情况可用积分牛顿方程来获得:

$$F_i(t) = m_i a_i(t) = m_i \frac{\partial^2 r_i(t)}{\partial t^2},$$

$$i = 1,2,\cdots,N(N\text{ 为分子的原子个数}) \qquad (2.6.2)$$

其中,$F_i(t)$为原子 i 在 t 时所受的力,m_i 为原子 i 的质量,$a_i(t)$为原子 i 在 t 时的加速度,$r_i(t)$为原子 i 在 t 时的位置。原子 i 所受的力可由分子力学势函数的负梯度求得:

$$F_i = -\frac{\partial}{\partial r_i}V(r_1,r_2,\cdots,r_N)。 \qquad (2.6.3)$$

3. 自由能微扰(FEP)方法

自由能是物理化学中的重要概念,它可用来描述分子间相互作用(缔

合)以及化学反应的趋势。将其应用于生命科学,则可计算生物大分子体系相互作用时的自由能变化趋势、酶催化反应的自由能变化、药物-受体结合时的亲和性以及溶剂效应等[9]。

根据统计力学,分子体系的自由能可由下式计算:

$$G = -k_B T \ln Z^N / V^N, \tag{2.6.4}$$

其中G是吉布斯自由能,Z是正则系综配分函数,k_B为玻尔兹曼常数:

$$Z^N = \int \cdots \int \exp[-\beta H(r_1, r_2, \cdots, r_N)] dr_1 \cdots dr_N,$$

$$\beta = 1/k_B T。 \tag{2.6.5}$$

对于非扩散体系,可以直接用几率分布函数计算自由能,对于生物大分子或溶液中的分子,由于其可到达的状态比较多,必须用其他方法计算,其中常用的方法为自由能微扰技术。状态2和状态1的自由能差ΔG为:

$$\Delta G = G_2 - G_1 = -k_B T \ln Z_2 / Z_1, \tag{2.6.6}$$

在计算时设置一联系两状态的耦合参数λ,哈密顿量H是λ的函数$H(\lambda)$。

在实际应用中一般用如下热力学循环计算自由能变化:

$$\begin{array}{ccccc} X+Y & \xrightarrow{\Delta G_1} & X:Y & \xrightarrow{\Delta G_5} & X:Y \\ \downarrow \Delta G_3 & & \downarrow \Delta G_4 & & \downarrow \Delta G_7 \\ X+Y' & \xrightarrow{\Delta G_2} & X:Y' & \xrightarrow{\Delta G_6} & X:Y' \end{array} \tag{2.6.7}$$

自由能计算的热力学循环可用于计算两个相似的配体(酶抑制剂、底物等)与同一生物大分子结合性的差异[9],这里X为生物大分子,Y和Y'为两个相似的配体,ΔG_1和ΔG_2为实验值,ΔG_3和ΔG_4可通过FEP方法求得,由热力学循环得:

$$\Delta\Delta G = \Delta G_2 - \Delta G_1 = \Delta G_4 - \Delta G_3, \tag{2.6.8}$$

通过这一公式可用纯计算的方法求$\Delta\Delta G$,由此推测配体Y和Y'与生物大分子结合性的差异。如果将X当作配体,Y当作生物大分子(如酶),Y上的部分残基定点突变后变为Y',这时上述热力学循环可计算生物大分子结构变化后对同一配体结合性的差异,如酶活性部位的残基定点突变后其催化活性的变化情况。在式(2.6.7)所示的热力学循环中,ΔG_5和ΔG_6

是实验值，ΔG_4 和 ΔG_7 可由 FEP 方法计算，即可由下式计算生物大分子化学反应（如酶催化反应）活化能的变化：

$$\Delta\Delta G' = \Delta G_6 - \Delta G_5 = \Delta G_7 - \Delta G_4。\tag{2.6.9}$$

同时热力学循环还可用于溶剂效应的计算[9]。

4.　量子力学计算方法

根据量子力学原理，分子的所有性质均可通过解薛定谔方程来获得[3]：

$$H\Psi = E\Psi,\tag{2.6.10}$$

其中 H 为哈密顿算符，Ψ 为波函数，E 为体系的能量或称为本征值。到目前为止，只求得氢分子离子薛定谔方程的精确解析解，其他分子体系只能以数值方法求其近似解，于是产生了各种不同的量子化学计算方法。

(1) 量子化学从头计算(ab initio)方法

ab initio 方法是在非相对论近似、核冻结近似（玻恩-奥本海默近似）和轨道近似（单电子近似）的基础上，将分子轨道表示成原子轨道的线性组合（AOLCMO），用自洽场方法解 Hartree - Forck - Roothaan(HFR)方程，得到体系的波函数，再计算体系的各种性质[31]。ab initio 方法的优点是比较严格，计算时除基函数外，不需任何经验参数，缺点是所能计算的体系较小。但随着计算机的发展和计算方法的改进，ab initio 方法所能计算的体系比 10 年前大为提高。

① Gaussian 92/DFT[6]能计算含有 100 个重原子的体系，目前所能计算的最大分子为有 6 对碱基组成的含有 387 个原子的 DNA 片段；

② SPARTAN[32]也能计算含有 100 个重原子的体系，所能计算的最大基函数为 600 个；

③ Turbermol[33]能计算的最大体系为 $C_{60}F_{60}$。

(2) 半经验量子化学计算方法

半经验量子化学计算方法的原理与 ab initio 计算方法相同，只是在求解 HFR 方程时，忽略一些双电子积分，在一些积分值的选择上使用实验值拟合的参数，在计算时只计算价电子，将内层电子加入到有效势能中去。这些方法虽然理论上不够严格，但计算速度比 ab initio 快 100 倍以上，最近发展了的一些半经验量子化学计算方法，如 MNDO、AM1 和 PM3 等，由于参数与实验值拟合得较好，其计算的结果可以和小基组 ab

initio计算结果相比[34]。

半经验量子化学计算方法所能计算的体系比 ab initio 大，目前几个常用的半经验量子化学计算程序所能计算的最大体系为：

① SPARTAN[32]含 200 个重原子的体系；

② MOPAC[35]含 200 个重原子的体系；

③ ZINDO[36]含 500 个金属原子的金属原子簇(cluster)体系。

(3) 密度泛函理论(density - functional theory, DFT)

近年来，在计算物理和计算化学中出现了一种非常活跃的方法——密度泛函理论(DFT)[37]，这一方法所能计算的体系比 ab initio 大，精度比半经验量子化学方法高。

早在 20 年代，Thomas[38]和 Fermi[39]就提出了原子的电子气模型(T - F模型)，指出电子均匀地分布在六维相空间中，以 $2e/h^3$ 的流量运动。根据这一假设，他们得到了电子运动的动能公式：

$$T_{TF}[\rho] = C_F\int\rho^{5/3}(r)\mathrm{d}r,\quad C_F = \frac{3}{10}(3\pi^2)^{2/3} = 2.871, \tag{2.6.11}$$

能量为：

$$E_{FT} = C_F\int\rho^{5/3}(r)\mathrm{d}r - Z\int\frac{\rho(r)}{r}\mathrm{d}r + \frac{1}{2}\iint\frac{\rho(r_1)\rho(r_2)}{|r_1 - r_2|}\mathrm{d}r_1\mathrm{d}r_2。\tag{2.6.12}$$

在 T - F 模型下，体系的能量是其电子密度的泛函，

$$E = E[\rho(r)]。\tag{2.6.13}$$

无论是从头计算还是半经验量子化学计算方法中，由哈密顿量 H 描述的电子体系，能量总是波函数的泛函。用变分原理和自洽场方法可求波函数和能量及体系的其他性质。而 T - F 电子气模型将能量当作电子密度的泛函，用变分法和自洽场方法求电子密度 $\rho(r)$ 和体系的能量 $E[\rho(r)]$，这一途径称为密度泛函理论(DFT)。

Hohenberg 和 Kohn[40]在 1964 年提出了 Hohenberg - Kohn 定理，可以说是真正 DFT 的开端。

Hohenberg - Kohn 第一定理：在外场 $V(r)$ 中 N 个相互作用电子体系的能量是由电子密度 $\rho(r)$ 决定的，即能量是 $\rho(r)$ 的泛函，

$$E = E[\rho(r)]。\tag{2.6.14}$$

这时，体系的能力可表示为：

$$E_V[\rho] = T[\rho] + V_{ne}[\rho] + V_{ee}[\rho] = \int \rho(r)V(r)\mathrm{d}r + F_{TF}[\rho]。 \quad (2.6.15)$$

其中

$$F_{TF}[\rho] = T[\rho] + V_{ee}[\rho]。 \quad (2.6.16)$$

将 $V_{ee}[\rho]$拆成两项

$$V_{ee}[\rho] = J_{ee}[\rho] + E_{XC}[\rho], \quad (2.6.17)$$

其中

$$J_{ee}[\rho] = \frac{1}{2}\iint \frac{1}{r_{12}}\rho(r_1)\rho(r_2)\mathrm{d}r_1\mathrm{d}r_2。 \quad (2.6.18)$$

$E_{XC}[\rho]$是电子交换相关(exchange - correlation)能，它是 DFT 计算的关键。

Hohenberg - Kohn 第二定理：$\tilde{\rho}(r)$为尝试电子密度，并且 $\tilde{\rho}(r) > 0$ 及 $\int \tilde{\rho}(r)\mathrm{d}r = N$，则有

$$E_0 \leqslant E_V(\tilde{\rho})。 \quad (2.6.19)$$

这里 $E_V[\tilde{\rho}]$的形式同式(2.6.14)，这是 DFT 的能量最小原理，也即密度函数变分原理。Hohenberg - Kohn 定理告诉我们由 $\rho(r)$计算基态分子的性质的基本原理，但没有告诉我们在不知道波函数的情况下怎样由 $\rho(r)$ 计算 E_0。这一问题由 Kohn 和 Sham 解决。将式(2.6.16)，式(2.6.17)代入(2.6.15)得：

$$E[\rho] = \int \rho(r)V(r)\mathrm{d}r + T[\rho] + J_{ee}[\rho] + E_{XC}[\rho]。 \quad (2.6.20)$$

与 HF 方法类似，Kohn 和 Sham 提出用下式代替(2.6.20)式：

$$E[\rho] = -\frac{1}{2}\sum_{i=1}^{n}\langle \varphi_i(1)|\nabla_i^2|\varphi_i(1)\rangle - \sum_A \int \frac{Z_A\rho(1)}{r_{1A}}\mathrm{d}r_{1A} + \frac{1}{2}\iint \frac{\rho(1)\rho(2)}{r_{12}}\mathrm{d}r_1\mathrm{d}r_2 + E_{XC}[\rho], \quad (2.6.21)$$

其中$\varphi_i(1)$，$i=1,2,\cdots,n$ 为 Kohn - Sham 轨道，同时他们还提出：

$$\rho = \sum_{i=1}^{n}|\varphi_i|^2。 \quad (2.6.22)$$

运用变分方法得 Kohn - Sham 方程：

$$F_{KS}(1)\varphi_i(1)=\varepsilon_{iKS}\varphi_i(1)。\tag{2.6.23}$$

其中，F_{KS}为 Kohn－Sham 算符。

$$F_{KS}=-\frac{1}{2}\nabla_i^2-\sum_A\frac{Z_A}{r_{1A}}+J(1)+V_{XC},\tag{2.6.24}$$

$$J(1)=\sum_j\int|\varphi_j(2)|^2\frac{1}{r_{12}}\mathrm{d}r_2,\tag{2.6.25}$$

$$V_{XC}=\frac{\delta E_{XC}[\rho]}{\delta\rho}。\tag{2.6.26}$$

V_{XC}为交换相关势。

F_{KS}与 Fock 算符非常类似，只是其中的交换算符被 $V_{XC}(1)$代替。

Kohn－Sham 轨道 $\varphi_i(1)$没有任何物理意义，DF 分子波函数也不是自旋轨道的 Slater 行列式，事实上并不存在 DF 分子波函数。Kohn－Sham 轨道能也不同于分子轨道能。

上已提及，DFT 的关键是计算 $E_{XC}[\rho]$，如知道了 $E_{XC}[\rho]$，一切问题均迎刃而解。围绕 $E_{XC}[\rho]$，提出了不同的近似方法，定域密度近似(LDA)即为其中行之有效的近似方法(详见文献[37]第8章)。

(4) *ab initio* 赝势法

在10年前，*ab initio* 赝势法还只能计算含几个重原子的体系，如今这一方法发展很快，已能计算上千个原子的体系[12]。*ab initio* 赝势法发展快的原因：

① *ab initio* 赝势法与 DFT 相结合[42]；

② Car 和 Parrinello 发展了动力学计算方法[43]；

③ Teter 等发展了共轭梯度法，用模拟退火法代替求二阶导数的自洽场迭代方法，在能量优化时还用了傅里叶变换技术[44,45]；

④ 并行计算机和并行算法应用于 *ab initio* 赝势法[42]。

目前 *ab initio* 赝势法还只是应用于材料科学，由于生物分子的对称性没有材料分子高，所以 *ab initio* 赝势法所能计算的生物体系还不大，但其方法和算法应可借鉴。

5. 量子力学(QM)和分子力学(MM)或分子动力学(MD)相结合的计算方法(QM/MM 模型)

在研究生物大分子体系的相互作用时，有时要牵涉到化学键的生成

和断裂，如酶催化反应机理的理论研究。分子力学方法虽然在近几年来也在发展过渡态结构的优化方法，但总的来说，酶催化反应机理的理论研究还是要用量子力学方法来计算。上已谈到，QM 所能计算的体系比较小，像对酶这样的大分子，目前的计算水平还不能实现。因此，最近发展了量子力学(QM)和分子力学(MM)或分子动力学(MD)相结合的计算方法(QM/MM 模型)[9,10]。QM/MM 模型的基本思想是：将所要研究的体系分成两部分，即量子力学部分(QM)和分子力学部分(MM)，在 QM 和 MM 之间有一相互作用部分(QM/MM)(见图 2.6.1)。在计算过程中将生物大分子的活性部位用量子力学计算，而将生物大分子的其他部位(包括溶剂)用分子力学、分子动力学或 FEP 方法计算。据此，可将体系的有效哈密顿算符 H_{eff} 分成以下几个部分：

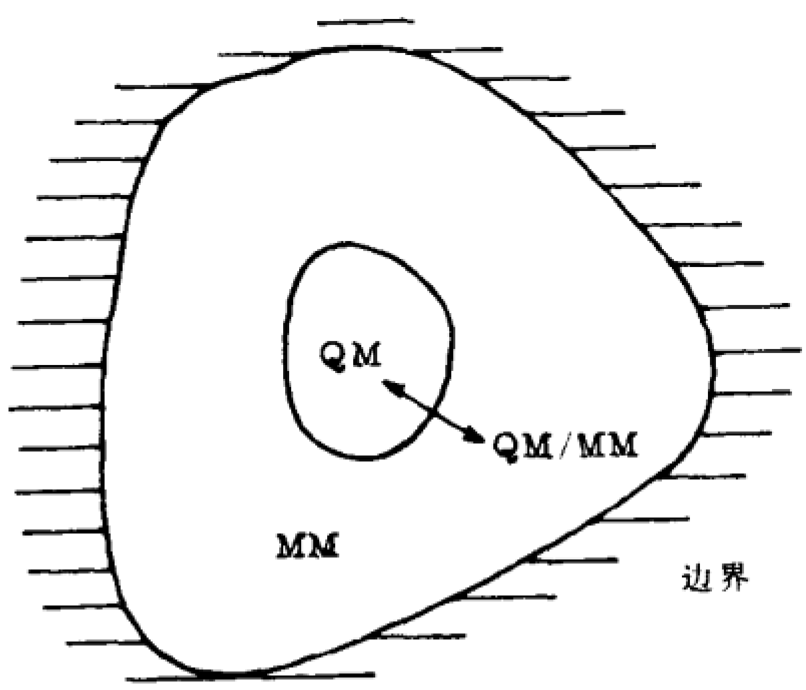


图 2.6.1 QM/MM 计算模型示图

$$H_{eff} = H_{QM} + H_{QM/MM} + H_{MM} + H_{边界}, \tag{2.6.27}$$

其中 H_{QM} 是体系 QM 部分的哈密顿算符；H_{MM} 是体系 MM 部分的相互作用能，用经典力场计算；$H_{QM/MM}$ 是 QM 部分和 MM 部分的相互作用能，由下式决定，

$$\begin{aligned} H_{QM/MM} &= H_{QM/MM}^{elec} + E_{vdw} \\ &= -\sum_{s=1}^{S}\sum_{i=1}^{2N}\frac{eq_s}{R_{si}} + \sum_{s=1}^{S}\sum_{m=1}^{M}\frac{q_s Z_m}{R_{sm}} \\ &\quad + \sum_{s=1}^{S}\sum_{m=1}^{M} 4\varepsilon_{sm}\left[\left(\frac{D_{sm}}{R_{sm}}\right)^{12} - \left(\frac{\sigma_{sm}}{R_{sm}}\right)^{6}\right], \end{aligned} \tag{2.6.28}$$

其中 e 是电子电荷，q_s 和 Z_m 分别是 QM 部分和 MM 部分的核电荷，S 和 M 分别表示 QM 部分和 MM 部分相互作用的原子核数，R_{si} 和 R_{sm} 分别表示 QM 部分的电子和核到 MM 部分原子核的距离。(2.6.28)式中的范德瓦耳斯作用能采用了 Lennard - Jones 模型，D_{sm} 和 σ_{sm} 是根据实验结果拟合的参数，它们由下式决定：

$$D_{sm} = (D_s D_m)^{1/2}, \quad \sigma_{sm} = (\sigma_s \sigma_m)^{1/2}, \tag{2.6.29}$$

这时

$$E = E_{QM} + E_{MM} + E_{QM/MM} + E_{边界}, \tag{2.6.30}$$

$$E = \langle \Psi | H_{QM} + H_{QM/MM} + H_{边界/QM} | \Psi \rangle / \langle \Psi / \Psi \rangle + E_{MM} + E_{边界/MM}。 \tag{2.6.31}$$

具体计算时，先用分子力学(MM)或分子动力学(MD)建立生物大分子的宏观系综，然后用自洽场方法分别解波动方程(2.6.32)。

$$H_{eff}\Psi = E\Psi。 \tag{2.6.32}$$

QM/MM 模型被广泛地应用于酶催化机理的理论研究、溶剂效应的理论计算以及其他配体-受体的相互作用[9,10,46]。

6. 经验价键理论(EVBT)

价键理论(VBT)是 Heitler 和 London 在处理 H_2 时形成的[3,46]。价键理论在处理分子时，将分子当做由原子壳层和成键价电子组成，其中的原子壳层由原子核和内实电子组成，价电子定域于原子壳层之间形成化学键，与 MO 相比，VBT 更接近于化学家的分子概念，物理意义较为明确，这一理论后来得到 Slater 和 Pauling 的发展，所以也叫做 HLSP 理论，我们以后称之为 HLSP - VBT，或简称为 VBT。VBT 理论提出后，得到了化学界的广泛应用，在 60 年代以前，在理论化学界占据主导地位。在 60 年代以后，由于 MO 理论的出现，特别是 MO 对称守恒定理以及前线轨道理论在解释化学反应规律方面取得极大的成功以后[31]，MO 在理论化学上的地位超过了 VBT。近年来，VBT 理论又受到了人们的重视，特别是 Warshel 等[3,46]将 VBT 与分子力学和分子动力学(MD)结合发展了经验价键理论(EVBT)。

上面我们已经讨论了 QM/MM 模型，这种模型是以分子轨道(MO)理论为基础的。与上述(QM/MM)模型不同，EVB 是以价键理论(VBT)为基础的，这是 Warshel 等[3,46]为研究溶剂效应以及酶催化反应而提出

来的一种半经验理论方法，从本质上来说，EVB 是 QM/MM 模型的一种，计算方法与 QM/MM 类似，只是将 QM 部分用 VB 理论计算。

7. SIBFA(sum of interaction between fragments computed ab initio)计算方法

SIBFA 法是又一种量子力学和分子力学相结合的计算方法，这一方法是由 Pulman 等发展的[47]。其基本方法为：

① 把生物大分子和药物分子等按其结构分成若干个片断，分别进行量子化学从头(ab initio)计算；

② 把从头计算得到的分子波函数转变成以原子和原子之间连线中点的多中心多极展开(multicenter multipolar expansion)，用来代替分子波函数，表征电子在整个分子空间的分布，用于计算各项相互作用能；

③ 把分子片断组成完整的分子，计算药物分子和生物大分子各自的构象能以及它们的相互作用能，对总体能量进行优化，得到药物-生物大分子复合物的最优构型。

在 SIBFA 计算时，构象变化引起的分子内部相互作用能的改变，即为构象能变化 ΔE_{conf}，由下列公式计算：

$$\Delta E_{\text{conf}} = E_{\text{MTP}} + E_{\text{pol}} + E_{\text{rep}} + E_{\text{disp}} + E_{\text{tor}}, \qquad (2.6.33)$$

其中 E_{MTP} 静电能，E_{pol} 为极化能，E_{rep} 为排斥能，E_{disp} 为色散能，E_{tor} 为扭曲能。分子间(药物和生物大分子之间)的相互作用能 ΔE_{inter} 用下列公式计算：

$$\Delta E_{\text{inter}} = E_{\text{MTP}} + E_{\text{pol}} + E_{\text{rep}} + E_{\text{disp}} + E_{\text{CT}}, \qquad (2.6.34)$$

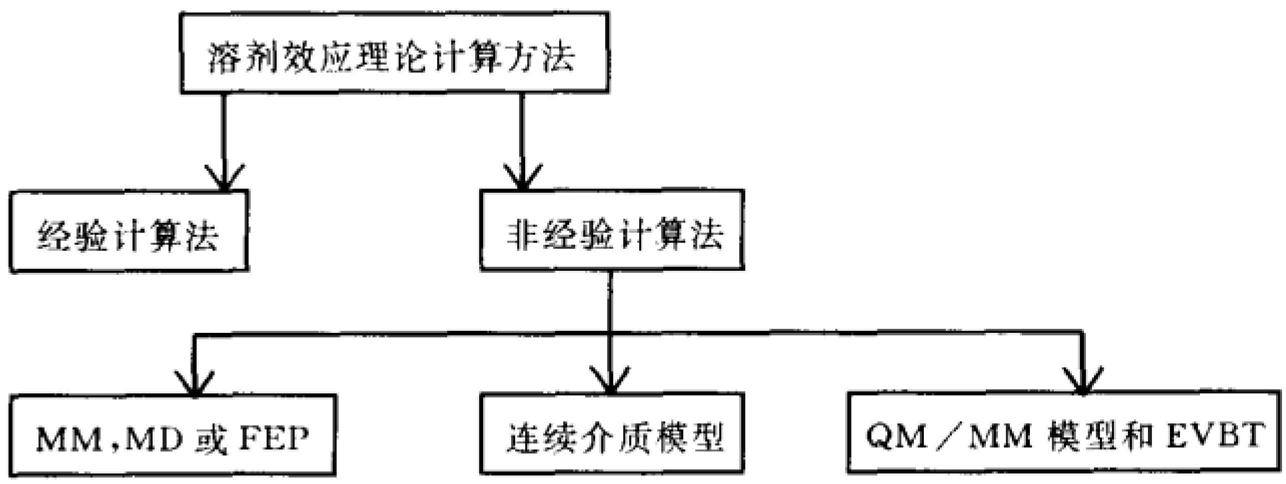
其中，E_{MTP}、E_{pol}、E_{rep} 和 E_{disp} 的含义同上，E_{CT} 为电荷转移能。

所以，如要分类的话，QM/MM 模型有三种，即 MO/MM、EVBT 和 SIBFA。

8. 环境的影响——溶剂效应的理论计算方法

化学是研究分子性质和反应的科学，许多化学过程是在溶液中进行的。溶剂对分子的构象、电子结构和化学反应有着重要的影响[3,46]。同时，溶剂对生物大分子的结构与功能也有重要影响，如溶剂能影响酶的催化活性、蛋白质的折叠和展开、生物大分子在溶液中的构象等，药物-受体相互作用中，疏水作用极为重要，因此对溶剂效应进行理论

研究有着重要的意义，它更能反映出溶液中分子的真实行为。所以近年来，发展了各种不同的理论计算方法。归结起来溶剂效应的理论计算方法如下框图所示：



(1) 经验计算方法

经验计算方法的实质是比较分子在气态和溶液态中的热力学性质，寻找纯态（气态）分子热力学性质与溶液中分子热力学性质（如溶解熵、溶解焓等）的关系或组成分子的基团或碎片与溶液中分子热力学性质的关系——经验公式[48]。对于有机溶剂，溶剂效应存在一种单一的线性关系(Barclay，Butler)[49]，并且不同溶质的经验公式基本相同。对于水溶液，虽然也存在经验关系，但不同的溶质有不同的经验公式，没有共性。所以水溶液的溶剂效应经验计算较为复杂[48]。

溶剂效应的纯经验计算方法基于假设：溶解性质是组成分子结构单元性质的总和。这一结论得到许多实验的证实。这一途径的缺点是不能在不同体系之间进行推广。这种方法构成基团贡献模型[48]。另一种基团贡献模型是：将溶解热力学性质 X_h^0 看作是“空穴”形成项的加和，“空穴”形成项是溶质分子与水分子之间的亲水作用（Y -）和疏水作用（R -）之和[47]。其线性关系为：

$$X_h^0 = X_{cav} + X_h^0(R-) + X_h^0(Y-) + \delta X_h^0(y_1, y_2, \gamma_{y_1 y_2})。 \quad (2.6.35)$$

这一模型的缺点是不能预测新分子的溶剂效应。

对于生物大分子，溶剂对其结构和功能的影响很大。Nozaki 等[50,51]将疏水能视为与溶剂非极化表面成正比，再用加和性计算溶剂总的疏水场分布。这一模型过于简单，为了克服这一缺点，有人将生物大分子分成

碎片并加以参数化，用溶剂可及性表面或用溶剂层体积来表示溶剂效应[48]，后者比前者更为精确。以溶剂层体积来表示溶剂效应的模型是用水化自由能来表示溶剂和溶质的相互作用，计算中使用的是经验力场。

(2) 连续介质模型

理论性更强、计算又较简单的模型是连续介质模型，将溶剂当作介电常数为 ε 的连续介质。Onsager 提出了反应场理论(reaction field theory)。在这个模型中，溶质被放在一个沉浸在介电常数为 ε 连续介质中的空穴内，这个空穴可以是球形、椭球形或者其他形状，为计算方便起见，通常选为球形(如图 2.6.2)。Tapia 和 Gosciaski 等将 Onsager 的反应场理论加入到量子化学分子轨道理论中，发展成为自洽反应场(SCRF)理论。此后，这一模型被广泛地应用于溶剂效应的研究中[52]。这一理论的计算方法大致如下。

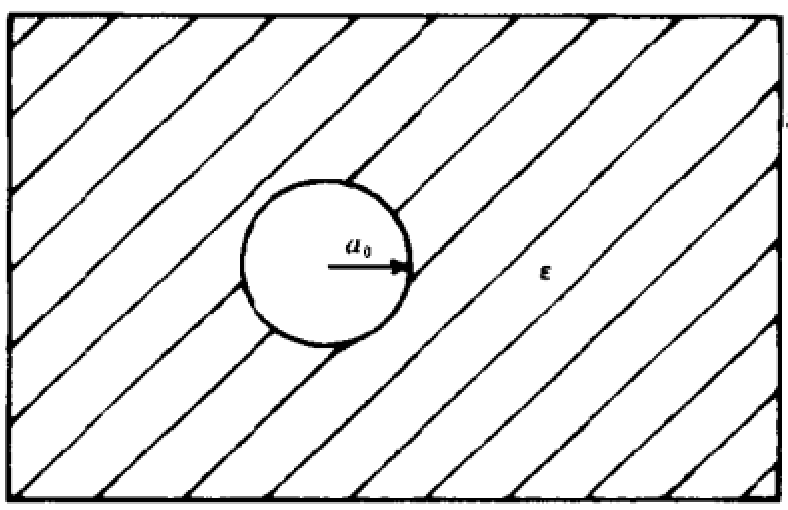


图 2.6.2 连续介质模型示意图

a_0 为溶质空穴半径，ε 为介电常数。

溶质分子固有的偶极矩 μ 作用于溶剂分子，使得溶剂分子产生一诱导偶极矩，而溶剂分子的诱导偶极矩反作用于溶质分子，从而对溶质分子产生一个额外的稳定作用。在分子轨道(MO)理论中，溶剂效应被当作溶质分子原有哈密顿算符 H^0 的微扰项 H'。若用 H^{rf} 表示溶质分子在连续介质中的哈密顿算符的话，则

$$H^{rf} = H^0 + H', \tag{2.6.36}$$

微扰项可以表达为溶质分子偶极矩算符 μ 和反应场 R 的耦合：

$$H' = -\mu R, \tag{2.6.37}$$

并且反应场 R 正比于分子的偶极矩 μ：

$$R = \mu g, \tag{2.6.38}$$

比例系数 g 反映反应场的强度，它与介质的介电常数和空穴半径 a_0 有关，其关系式如下：

$$g = 2(\varepsilon - 1)/(2\varepsilon + 1)a_0^3。 \tag{2.6.39}$$

在自洽场波函数中，反应场的溶剂效应被当作 Fock 矩阵的附加项进行计算：

$$F_{\lambda\sigma} = F_{\lambda\sigma}^0 - g\mu\langle\phi_\lambda|\mu|\phi_\sigma\rangle, \tag{2.6.40}$$

这里 ϕ_λ 和 ϕ_σ 为基函数。当溶剂的极化作用引入之后，分子的总能量可表达为：

$$E = \langle\Phi|H^0|\Phi\rangle - 0.5\mu R, \tag{2.6.41}$$

其中 Φ 是分子的总波函数。自洽反应场的计算过程如下：

① 选择一个反应场 R；

② 计算自洽场能量和偶极矩。第一次计算时将分子当作气态分子计算；

③ 计算新的反应场 $R=\mu g$，如不收敛，则返回到第二步直到自洽为止。

空穴半径选择有多种方法[52]，最简单的方法是根据分子的体积计算 a_0，

$$a_0^3 = 3V_m/4\pi N, \tag{2.6.42}$$

其中 V_m 是分子体积，由分子量除以密度直接得到，N 是阿伏伽德罗常数。具体计算时，将由(2.6.42)式计算的 a_0 加 0.05 nm，这样更接近实际情况[52]。

Stewart 将 SCRF 理论加入到半经验量子化学程序 AMI 中，编制了可计算溶剂效应的半经验量子化学程序 MOPAC。Wiberg 等将 SCRF 理论加到了 ab initio 程序 Gaussian 体系中，使得 Gaussian 程序新版本 Gaussian92 和 Gaussian94 可以计算溶剂效应[52]。后来 Sinanoglu 等[53]提出了"溶剂效应"理论，将溶剂化能分成溶质-溶剂作用能、空穴生成能和溶质在溶剂环境中的体积自由能，Onsager 的反应场只是其中的静电作用部分。"溶剂效应"理论比 Onsager 反应场理论更精确。另一种改进方法是用多极相互作用代替偶极相互作用，同时根据溶质分子的实际形状决定空穴的形状。

(3) 虚拟电荷模型(VCM)

VCM 以一系列虚设电荷代替分布在溶质周围的溶剂，用其来模拟溶剂的作用[54]，这时，溶质分子的哈密顿算符 H 为：

$$H = H_0 - \frac{\varepsilon - 1}{2\varepsilon}\left[\sum_{s=1}^{n}\sum_{i=1}^{m}\frac{Q_s}{r_{si}} - \sum_{s=1}^{n}\sum_{i=1}^{m}\frac{Q_s Z_i}{r_{si}}\right], \quad (2.6.43)$$

其中的 ε 为介电常数，括号中的第一项为溶质的电子与溶剂的相互作用能，第二项为溶质的核与溶剂的作用能。然后用与 SCRF 相同的方法解 HFR 方程。这一方法的缺点也是没有考虑溶剂之间的相互作用，在计算大体系时较费时。

最近，Still[55]提出了一种新的计算溶剂效应的模型，Cramer 和 Truhlar[56~58]将其加入到了半经验量化程序 AMI 中，编制了计算溶剂效应的程序 AMSOV。这一方法考虑的因素较多，他们将溶剂化自由能分成：

$$G_s^0 = G_{\rm ENP} + G_{\rm CD}^0, \quad (2.6.44)$$

$$G_{\rm ENP} = E_{\rm SCF} + E_{\rm NN} + G_{\rm P}, \quad (2.6.45)$$

其中 $E_{\rm SCF}$为溶质分子电子的 SCF 能，$E_{\rm NN}$为溶质分子的核排斥能，$G_{\rm P}$为溶质-溶剂相互作用的极化能，$G_{\rm CD}^0$是溶质分子空穴化能和扩散能。

$$G_{\rm CD}^0 = \sum_{k=1}^{N}\sigma_k A_k, \quad (2.6.46)$$

N 是溶质分子的原子数，A_k是溶质分子中原子 k 的溶剂可及性表面，σ_k是参数，成为溶剂可及性表面张量。$G_{\rm P}$由玻恩方程计算：

$$G_{\rm P} = -\frac{1}{2}\left(1 - \frac{1}{\varepsilon}\right)\sum_{k=1}^{N}\sum_{k'=1}^{N} q_k q_{k'}\gamma_{kk'}, \quad (2.6.47)$$

ε 是介电常数，$\gamma_{kk'}$是库仑积分，α_k是有效原子半径。Still 提出计算 $\gamma_{kk'}$的半经验公式：

$$\gamma_{kk'} = \frac{1}{\sqrt{r_{kk'}^2 + \alpha_k\alpha_{k'}\left[\exp(-r_{kk'}^2/d_{kk'}\alpha_k\alpha_{k'}) + c_{kk'}r_{kk'}\right]}}。 \quad (2.6.48)$$

$r_{kk'}$为原子 k 和 k'间的距离，$d_{kk'}$是常数。将上式代入 HFR 方程得：

$$G_s^0 = \frac{1}{2}\sum_{\mu=1}^{m}\sum_{\nu=1}^{m}P_{\mu\nu}[H_{\mu\nu} + F_{\mu\nu}^{(0)}] + \sum_{k=1}^{N}\sum_{k'=1}^{N}E_{kk'}, \quad (2.6.49)$$

其中

$$E_{kk'} = \frac{1}{2}\left[Z_k Z_{k'} g_{kk'} - \left(1 - \frac{1}{\varepsilon}\right)q_k q_{k'}\gamma_{kk'}\right] + \delta_{kk'}\sigma_{k'}A_k, \quad (2.6.50)$$

$$g_{kk'} = \begin{cases} 1/r_{kk'}, & k \neq k'; \\ 0, & k = k。\end{cases} \quad (2.6.51)$$

μ,ν 是价轨道指数，m 是价轨道数。

因

$$F_{\mu\nu} = \partial G_s / \partial P_{\mu\nu}, \tag{2.6.52}$$

得

$$F_{\mu\nu} = F_{\mu\nu}^{(0)} + \delta_{\mu\nu}\left(1 - \frac{1}{\varepsilon}\right) \sum_{k} \sum_{\mu' \subset k'} (Z_k - P_{\mu'\mu'}) \gamma_{kk'},$$

$$\mu \subset k, \tag{2.6.53}$$

$$\Delta G_s^0 = G_s^0 - E_{\mathrm{EN}}(g) + \Delta G_{\mathrm{vib}} + \Delta G_{\mathrm{ele}}。 \tag{2.6.54}$$

这样便可用自洽场方法解 HFR 方程。再用(2.6.54)式求溶剂化能，其中 g 代表气相分子，s 代表溶剂，Δ 代表溶液相与气相之差，$E_{\mathrm{NG}}(g)$ 为气相溶质分子的电子和核排斥能，ΔG_{vib} 为溶液分子与气相分子振动自由能之差，ΔG_{ele} 表示两相激发自由能之差。

Cramer 和 Truhlar 对 C、N、O、P、S 等元素在水溶液中的溶剂化参数作了优化，最近对这些元素在环己烷中的溶剂化参数也作了优化[57,58]。这一方法的优点是比较精确，物理意义也较明确，缺点是通用性不好、每一种溶剂都必须有一套参数。但因生物分子大多存在于体液中，而体液的主要成分是水，所以这一方法可用于生物分子溶剂效应的计算。

(4) 超分子模型(SMM)计算法

SMM 计算方法是在溶质分子周围设置一些溶剂分子，将溶质-溶剂体系当作一个分子(超分子)，用 MM、MD、FEP 或 QM 方法计算。其中 QM 方法可以是 ab initio 或半经验量子化学计算方法，用得较为成功的是 Pullman 等发展的定域轨道微扰组态作用(PCILO)方法[59,60]。其计算步骤为：

① 根据化学经验或分子静电势计算结果决定溶质分子与第一层溶剂结合的可能部位(通常是易形成氢键或与溶剂间有静电作用的部位)；

② 设置溶质与第一层溶剂作用的方式(溶剂分子的趋向)；

③ 设置第二层和更高层的溶剂分布；

④ 计算上述超分子的结构和性质。

这一方法的优点是，在计算过程中同时考虑了溶质-溶剂、溶剂-溶剂的相互作用，并且可以计算溶质-溶剂的局部作用(如氢键作用和电荷转移)。基于此，这一方法特别适合于计算水化作用。

(5) FEP 方法

用 FEP 方法计算溶剂效应，只要将热力学循环改为如下形式：[3,9,46]

$$\begin{array}{ccc} X(g) + Y(g) & \xrightarrow{\Delta G_1} & X:Y(g) \\ \downarrow \Delta G_3 & & \downarrow \Delta G_4 \\ X(S) + Y(S) & \xrightarrow{\Delta G_2} & X:Y(S) \end{array}, \tag{2.6.55}$$

$$\Delta G_{sol} = \Delta G_2 - \Delta G_1 = \Delta G_4 - \Delta G_3。 \tag{2.6.56}$$

(6) QM/MM 模型和 EVBT 计算方法

与生物大分子体系的 QM/MM 模型计算方法相同[10]，只是将溶质分子用 QM 方法计算，溶剂分子用 MM、MD 或 FEP 方法计算。EVB 计算方法前已提及，用价键理论计算溶质分子，用 MM、MD 或 FEP 方法计算溶剂分子[3,46]。

2.6.2 计算机辅助药物设计的基本方法和策略

计算机辅助药物设计是一门新兴边缘学科。它以计算机为工具，根据前人积累的有关生物活性物质结构与功能的资料，以及药物在体内作用"靶点"即所谓受体的三维结构知识，设计出具有特定疗效的药物分子，来进行有机合成和药理测试，从中发现高效、低毒并且有较好生物利用度的分子，进入新药研究的下一阶段。其特点是可以大大加快新药发现的速度，提高新药开发的效率，节约大量的人力、物力和财力[5]。根据目前国内外药物分子设计的研究状况，计算机辅助药物分子设计的方法和策略可由图 2.6.3 来表示。

由图 2.6.3 可知，计算机辅助药物设计一般有三条途径。一是在受体三维结构未知的情况下，从一系列具有生物活性的小分子开始，用量子力学(QM)、分子力学(MM)或分子动力学(MD)方法，对这些小分子进行构象分析，得到 QSAR 模型或药效基团模型，以及假想受体结构模型，然后对现有的化合物进行结构改造；或者根据上述 QSAR 模型和假想受体模型，用全新药物设计(de novo)方法设计新型结构的化合物，或在小分子结构数据库中搜寻出符合药效基团模型的化合物，然后对新设计的化合物进行有机合成，并用药理方法测试这些化合物的活性。经过这样的反复循环，选出有苗头的化合物进入临床前研究[5]。

第二条途径，是在受体生物大分子三维结构已经阐明的情况下，用数

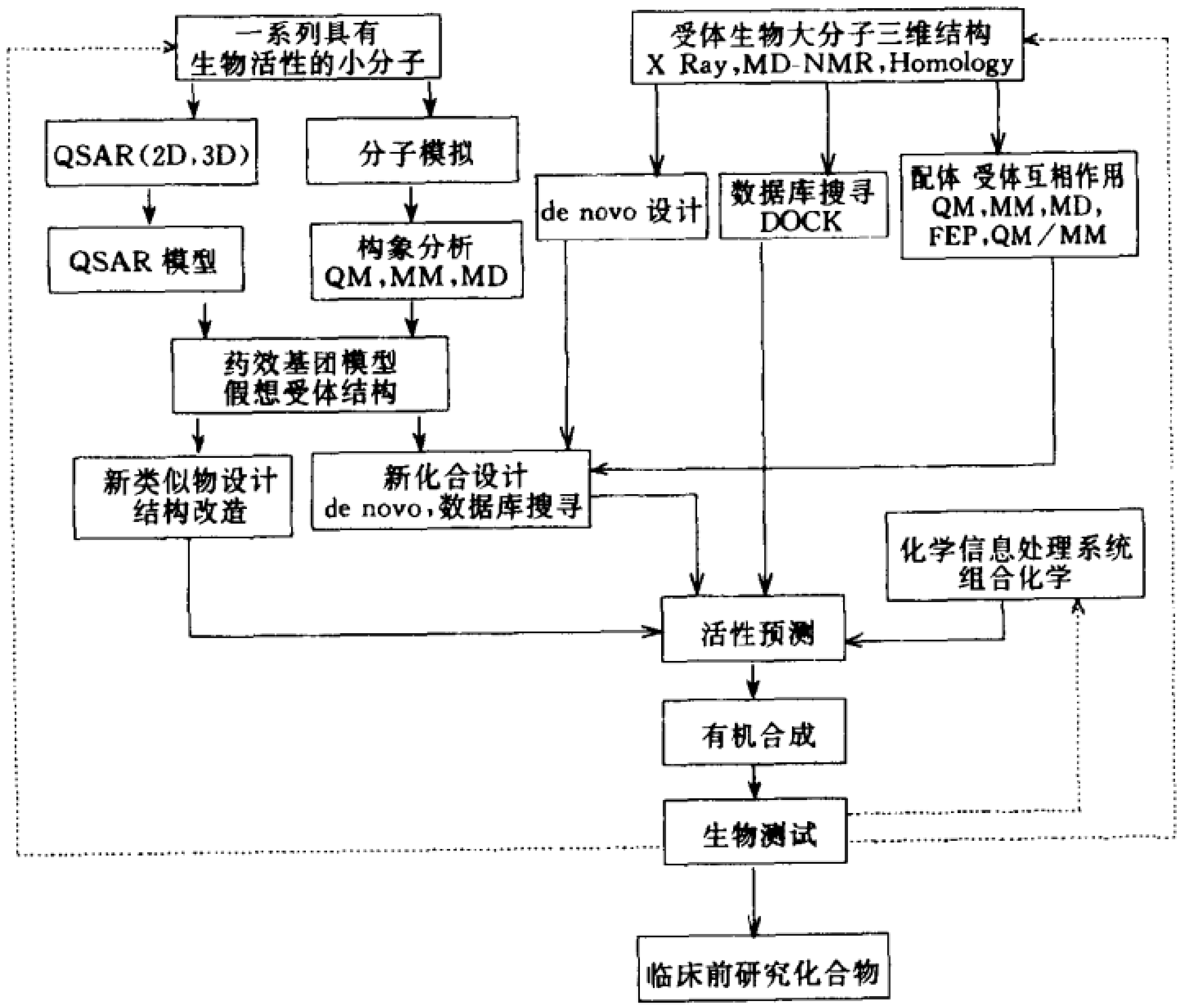


图 2.6.3 计算机辅助药物分子设计示意图

据库搜寻的方法,在三维结构数据库(如剑桥库,MDL 数据库)中搜寻几何形状、化学结构都与受体活性部位相互匹配的化合物,或者用全新药物设计的方法设计出几何形状和化学结构与受体活性部位相互匹配的结构,然后合成所设计的化合物,并进行药理测试。经过这样的反复循环,从中选出有苗头的化合物进入新药研究的下一阶段工作。其中受体三维结构可由 X 射线(X-Ray)、多维核磁共振(MD-NMR)或由同源蛋白建模(homology)方法获得[5]。

第三条途径是用组合化学方法[61],在定量构效关系、受体结构要求、分子量、分子的油水分布系数 $\log P$ 等约束条件下,建立化学反应生成物和产物结构库,然后根据分子的差异性等条件选出符合要求的化合物,并合成这些化合物去做药理测试,从中选出有前途的化合物进入临床前研究。

这三条途径是互相联系、相辅相成的。综合运用这些方法设计的新型结构化合物,还需进行有机合成和药理测试,所以,药物设计不仅是理论化

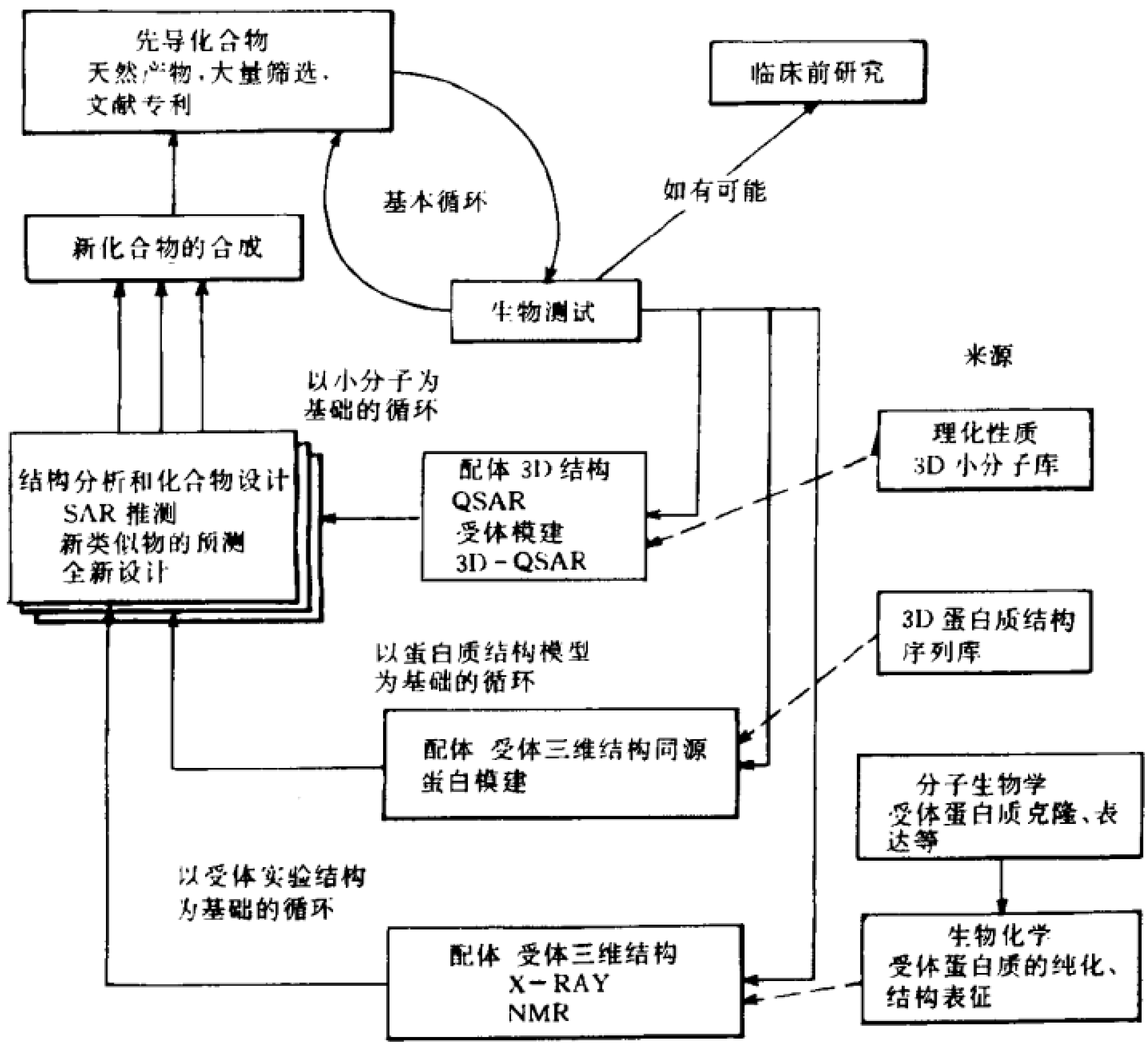


图 2.6.4　新药研究与开发示意图

学家和计算化学家的事，还必须与合成化学家、药理学家合作才能真正做到分子设计、有机合成和药理测试一条龙，经过多次循环，做出有实际意义的科研工作——发现新的先导化合物和高效、低毒、高生物利用度的药物。

药物设计是一门综合性的学科，从图 2.6.3 中可知，药物设计是计算化学、计算机科学、药理学和分子生物学的交叉学科。传统的发现新药方法（图 2.6.4 中基本循环）是用药理方法筛选大量人工合成或从天然产物中分离得到的化合物，有较大的盲目性，耗资大，周期长。计算机辅助方法加入到药物设计和发现新药的循环（图 2.6.4 中其他三个循环）后，可以大大加速新药发现的速度，降低新药开发的成本。计算机辅助药物设计的重要基础，就是生物大分子三维结构及其与配体相互作用的理论计算。

2.6.3 理论计算方法在生物体系研究和药物设计中的应用

1. 分子力学和分子动力学

(1) 药效基团的建立

在受体结构未知的情况下,可由分子力学方法来建立药物类似物的药效基团模型,为设计新化合物提供结构信息,这一技术是由 Marshall 等[62]提出来的。这一方法的基本思想是通过分子力学计算和构象分析,寻找出一系列化合物中共同的药效基团,并确定药效基团间的距离、键角和二面角,这样便得到了药效基团模型。确定药效基团模型后,根据化合物与药效基团迭合的好坏来判断这一化合物的生物活性的高低。

建立药效基团模型通常从研究刚性化合物的构象开始,因刚性分子与受体结合时,其骨架不易改变,所以刚性分子本身即是药物类似物活性构象的一个参照。如果化合物的柔性较大,则其全局极小构象并不一定是与受体结合时的活性构象,但可以肯定,活性构象一定是某一低能构象。所以柔性化合物活性构象的搜寻一般要以刚性分子的药效基团模型为模板,在限制条件下进行构象搜寻,找到满足药效基团结构的低能构象。

Hagler 等[63]用此方法研究了性腺激素释放荷尔蒙(GRH),对环状多肽拮抗剂与 GRH 结合时的药效基团模型进行了 24 ps 的动力学模拟,从其运动轨迹中每皮秒提取一瞬间结构用分子力学优化,所有的结构均落入同类的构型家族中,这一结构就被推断为 GRH 拮抗剂可能的与 GRH 结合时的活性构象。

应用分子力学和构象分析,可以得到柔性较高化合物的相应药效基团与药效基团模型迭合时的构象。但人们真正感兴趣的是在溶液环境中药物与受体生物大分子结合时的自由能。结合自由能可分成两部分,一部分是柔性类似物的构象限制(locking)在药效相关位置时所消耗的自由能 ΔG_{lock},即由柔性分子 D^{flex}变成活性构象 D^{lock}时所消耗的自由能,另一部分是活性构象与受体 R 结合时的自由能 ΔG_{bind}。式(2.6.57)示意了柔性分子和刚性分子与受体结合时的差异。因刚性分子本身即为活性构象,所以其 $\Delta G_{lock}=0$。

$$D^{flex} + R \xrightarrow{\Delta G_{lock}} D^{lock} + R \xrightarrow{\Delta G_{bind}} DR^{flex},$$

$$D^{rigid} + R \xrightarrow{\Delta G = 0} D^{rigid} + R \xrightarrow{\Delta G_{bind}} DR^{rigid}, \qquad (2.6.57)$$

其中的 DR 为药物-受体复合物，必须注意的是，上述柔性类似物与受体的作用方式必须与刚性化合物一致，这样它们与受体作用时的差别不太大。据此可以推测，柔性化合物和刚性化合物与受体结合亲和性的差异关键在于 ΔG_{lock}。由玻尔兹曼规律知，在构象系综中活性构象存在的概率 P_{lock}与 ΔG_{lock}的关系为：

$$\Delta G_{lock} = -RT\ln P_{lock}, \tag{2.6.58}$$

且 P_{lock}是可得到构象系综的函数，

$$P_{lock} \approx \frac{\exp(-E_{lock}/RT)}{\int \exp[-E(r)/RT]\mathrm{d}r} \approx \frac{\exp(-E_{lock}/RT)}{\sum_i \exp(-E_i/RT)}, \tag{2.6.59}$$

其中 E_{lock}为活性构象的能量，R 为气体常数，T 为绝对温度，积分近似为仅与每个能量极小值 E_i有关项的加和，上式的第二个假设是柔性构象和活性构象的溶剂化自由能相同。在这几种假设下，药物结合性的比较只能在非常相似的类似物之间进行，并且要保证每个类似物与受体作用的方式相同，因此上述关系只是一个半定量的规律。如果所选的刚性化合物是一较好的活性构象模板，且上述能量假设均成立，则柔性分子变成活性构象所消耗的自由能 ΔG_{lock}每 5.7 kJ/mol，其活性就会降至1/10，因此 ΔG_{lock}与活性的关系十分密切。

Andrea 等[64]对甲状腺素类似物进行了全系综构象分析，结果表明 ΔG_{lock}与体外活性的相关性为 $r=0.89$，与体内活性的相关性也较好。虽然他们在计算构象能时用的是半经验量子化学计算方法，但其结果表明，系综分析时应用自由能来代替内能 E，往往得到较好的计算结果。

对于一个旋转单键较多的分子，即使是分子力学，要计算出每个极小值的能量也是非常困难的，下面一个例子说明了 ΔG_{lock}可近似地用分子力学能量差表示。

Pettersson 和 Liljefors[65]用分子力学构象分析，搜寻了多巴胺受体激动剂 R -阿朴吗啡类似物的药效构象，R -阿朴吗啡是多巴胺受体的高效、高立体选择性激动剂，其药效基团由四部分组成(如图 2.6.5 所示)：两个羟基，一个苯环中心及氮原子，因阿朴吗啡是刚性分子，所以它的结构可以当作活性构象的模板。在进行其他柔性类似物构象搜寻时，加上上述阿朴吗啡药效基团的结构限制，使得每个类似物的构象的这四个点与

阿朴吗啡能最佳重叠，在这种情况下获得的最低能量构象即为此类似物的活性构象。活性构象与全局极小构象的能量差 ΔE_{lock} 可近似地认为是 ΔG_{lock}。Pettersson 和 Liljefors 等的计算结果表明：对于 7 个低活性和无活性的类似物，$\Delta E_{lock} \geqslant 10.5$ kJ/mol；2 个中等活性的类似物其 ΔE_{lock} 为 8.4～12.5 kJ/mol；对于 3 个高活性的类似物，$\Delta E_{lock} < 4.2$ kJ/mol。一般认为：如果 $\Delta E > 20$ kJ/mol，而药效基团模型正确的话，则此化合物通常是没有活性的。

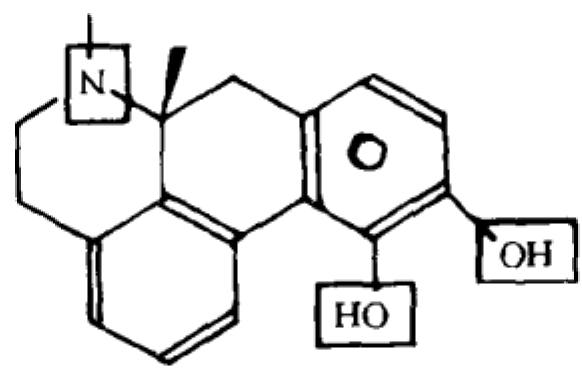


图 2.6.5 阿朴吗啡药效基团

（2）配体-受体相互作用

受体通常是一些生物大分子——蛋白质、核酸或多糖。随着分子生物学和 X 衍射晶体学、多维 NMR 等实验技术及计算机技术的不断发展，许多受体生物大分子的三维结构已经测定，有些蛋白质受体结构虽然未知，但可根据其与其他蛋白质的同源性，用同源蛋白搜寻等分子模拟方法建立它们的三维结构，这就为在受体三维结构的基础上研究配体-受体相互作用、进行药物设计奠定了基础。

在受体三维结构的基础上进行药物设计，首先必须将药物小分子结合到受体大分子的活性部位上去，这一过程称为分子对接(docking)。分子对接一般是根据人的经验，应用实时立体图形技术，在计算机屏幕上进行手工操作，选择各种可能的不同方向进行对接，然后经能量优化得到药物-受体复合物的结构。Kuntz 等发展了一种自动分子对接的算法，这一方法搜寻刚性配体与受体立体作用的方向，能重现已知配体的结合方向并确定一个新配体的结合方向。后来 Kuntz 等又将这一方法发展为可以部分考虑配体的柔性，并用这一方法在受体结构的基础上搜寻、筛选数据库，根据配体与受体相互作用的优劣打分。这可以认为是一种计算机筛选新药的方法。Kuntz 等用这方法发现了具有抗 HIV－1 和抗疟等作用的

先导化合物[66]。

一旦得到药物-受体复合物的结构之后,就可以优化药物及其周围活性部位的结构,计算药物-受体的相互作用能,从结果可以判断不同类似物与受体的结合能与实验活性值之间是否有相关性。如果有相关性,就可以用相互作用能来预测未知化合物的活性。

然而,用分子力学计算的相互作用能来预测化合物的活性是非常困难的,因为真正的药物-受体相互作用的能量关系必须用溶液中的 ΔG_{aq} 表示,即用自由能微扰和热力学循环来计算。

$$\begin{array}{ccccc} D(g) & + & R(g) & \xrightarrow{\Delta G_g} & DR(g) \\ \downarrow{\scriptstyle \Delta G(D)} & & \downarrow{\scriptstyle \Delta G(R)} & & \downarrow{\scriptstyle \Delta G(DR)} \\ D(aq) & + & R(aq) & \xrightarrow{\Delta G_{aq}} & DR(aq) \end{array} , \tag{2.6.60}$$

式(2.6.60)中的 aq 表示溶液,g 表示气相。分子力学计算的只是分子内能,而不是自由能,所以不能计算包含溶剂的作用。实际上分子力学计算的相互作用能是 ΔE_g。在比较相似度较高的类似物时,$\Delta\Delta E_g$可近似地认为等于 $\Delta\Delta G_g$。因此 Blaney 等[67]认为,如果比较两个药物,则

$$\Delta\Delta G_{aq} \approx \Delta\Delta E_g + \Delta\Delta G(D), \tag{2.6.61}$$

其中 $\Delta\Delta E_g$是两药物分子与受体相互作用的分子力学相互作用能之差,$\Delta\Delta G(D)$是两药物分子溶剂化自由能之差。Blaney 等[67]用此方法计算了甲状腺素类似物与前清蛋白相互作用自由能,相对的结合性与活性有很好的一致性。式(2.6.61)还被应用于 dioxin 类似物与前清蛋白[68]以及三甲氧苄二氨嘧啶类似物与二氢叶酸还原酶(DHFR)[69]结合能的计算。活性与结合能计算值有良好的相关性,并且还成功地预测了两个新的 DHFR 抑制剂。

Caldwell 和 Kollman[70]在 Patel[71]的 NMR NOE 数据的基础上,用分子力学优化了纺锤菌素与 d(CGCGAATTCGCG)$_2$复合物的结构,计算结果与晶体结构[53]符合,两者最小二乘重迭总偏差小于 0.12 nm,如不考虑端基 amidinum,则最小二乘重迭总偏差为 0.06 nm。

Dicapua 和 Swaminathan[72]用分子动力学方法模拟了 Gly_{30}和 Ala_{30} 在真空和溶液中的构象,计算结果表明,甘氨酸螺旋结构的弯曲是 Gly 组

成的多肽所固有的性质，而丙氨酸螺旋结构的弯曲是由于水插入分子内氢键的结果。在 250 ps、300 K 的动力学模拟中，没有发现水分子插入现象；在 30 ps、350 K 的动力学模拟时，发现了水插入现象。这一计算结果支持了 Sundaralingam 等[73]实验结果。水的插入是蛋白质展开的重要因素。Tirado - Rives 等[74]对核糖核酸 S 肽进行了 300 ps、780 K 和 500 ps、358 K 的动力学模拟，计算结果表明，低温时 α 螺旋结构是稳定的，高温时 α 螺旋氢键变成 3_{10}螺旋氢键结构，这也与实验结果一致。

总之，分子力学可应用于各种受体生物大分子-药物相互作用的研究中，但因分子力学计算的仅是分子内能，没有考虑环境和溶剂的作用，所以计算结果往往只是定性或半定量的。要进一步获得比较定量的计算结果，可用分子动力学方法和自由能微扰途径来计算，这两种方法均以分子力学力场为基础，因此广义地说，也是分子力学方法。

2. 自由能微扰(FEP)

(1) 蛋白质-配体相互作用

应用循环(2.6.7)可计算蛋白质-配体相互作用。应用这一方法研究蛋白质-配体相互作用的第一个例子是 McCammon 等[75]计算胰蛋白酶与其抑制剂苯甲脒类似物的相互作用，他们用 F 取代甲脒上的 H，计算的 $\Delta\Delta G$ 与实验结果一致。

Bash 等[76]及 Merz 和 Kollman[77]用 FEP 方法研究了嗜热菌蛋白酶(TMS)抑制剂 X - PO_2- Y - Z 与 TMS 的结合性，其中 X 和 Z 是大的疏水性基团，Y 是可变基团，Bash 等计算了 Y＝NH 和 Y＝O 时抑制剂与 TMS 的结合性，计算结果表明前者的结合性比后者好，$\Delta\Delta G$～16.8 kJ。Merz 计算了 Y＝CH_2的情况，计算结果表明，Y＝CH_2和 Y＝NH 的结合性相差不多，这一结果被后来的实验所证实[78]。

Merz 等[79]用 FEP 方法研究了碳酸酐酶(CA)抑制剂 n -己基氨磺酰点突变后与 CA 的结合性，他们将己基改变成 H，SO_2NH^- 改变成 SO_3^-，计算结果与实验值一致，并且 SO_2NH^- 改变成 SO_3^- 的计算结果与 pK_a差别非常吻合。

Menziani 等[80]也用 FEP 方法研究了 CA 抑制剂与 CA 的结合性，用 Cl 取代苯氨磺酰的 4 - H，$\Delta\Delta G$ 的计算值 －4.2 kJ/mol 与实验值 －6.3 kJ/mol非常吻合，后来，Hoops[81]等根据 MNDO 方法电荷分布计

算结果，将 Zn 离子的正电荷离域于配体组氨酸，然后进行 FEP 计算，计算结果比原先的 Zn^{2+} 模型要好一些。

Rao 和 Singh[82]对天冬酰胺蛋白酶与其抑制剂的作用进行了 FEP 研究，计算结果表明，天冬酰胺蛋白酶活性部位两个二重轴对称的天冬氨酸之间的质子转移的灵敏性，决定它们的几何构型。量子化学计算结果表明，质子转移反应的活化能很低，他们将天冬酰胺蛋白酶抑制剂胃素上的 OH 用 H 取代后变成去羟胃素 FEP；计算结果 $\Delta\Delta G$ 为 20.9 kJ/mol，与实验结果一致。

最近，不少人用 FEP 方法研究 HIV 蛋白酶抑制剂。Reddy 等[83]用 FEP 计算了 HIV 蛋白酶的两个肽类抑制剂的结合性，一个是五肽，一个是六肽，从六肽抑制剂中去掉一个缬氨酸残基后，结合自由能减少 13.8±4.6 kJ/mol，与实验值 15.88±5.43 kJ/mol 很接近。Reddy 等[84]还利用 FEP 计算结果进行新型 HIV 抑制剂的设计工作，设计结果由 Aguron 制药公司进一步研究。另外有两个独立的研究小组用不同的方法[85,86]，研究了 HIV 蛋白质羟基亚己基抑制剂 JG365 的 R/S 选择性。这两个研究小组在计算时用不同的方法将一个质子连接在 ASP-25/125 残基上，得到相同的结论，即 JG365 与 HIV 蛋白酶结合时，ASP-25/125 先质子化；同时 FEP 计算结果与实验结果也一致。Ferguson 等[85]还预测了去羟 JG365 的结合性。Cieplak 等[87]进一步研究了 JG365 抑制剂，将其中的酰胺基用亚乙基和氟代亚乙基等排体取代后，结合性均有明显提高。计算结果表明 JG365 结合性比突变产物低的原因是极性酰胺不能与蛋白质很好地作用，从而导致结合性较差。Cieplak 等还对 JG365 进行了结构改造，并计算了它们的结合性。

Aguron 制药公司将 FEP 计算方法用于腺苷合成酶(TS)抑制剂的设计上。Reddy 等计算的炔丙基类和醛类抑制剂的相对结合性及溶剂化自由能与实验结果完全一致。

Cummins 等[88]和 Brooks 等[89~92]用 FEP 方法对二氢叶酸还原酶(DHFR)的结合性进行了广泛研究。前者对辅助因子 NADP 的还原态及氧化态与 DHFR 结合的差异进行了 FEP 计算，结果表明，还原态双配体复合物(二个配体与一个生物大分子结合)的结合性比包含底物的四配体复合物(四个配体与一个受体结合)的结合性要强，并对这一现象从结构方面作了解

释。后者对三甲氧苯二氨嘧啶与 DHFR 的结合性进行了 FEP 研究。

FEP 计算方法不仅可以用来研究配体受体的结合性，还可用于酶的催化反应。Caldwell 等[93]用 FEP 研究了 α 溶解酶与底物作用的过渡态稳定性，所用的热力学循环如(2.6.7)式所示，X 代表酶，Y 和 Y′代表两个不同的底物，XY 和 XY′是两个 Michaelis 非键复合物。他们先用量子力学优化模型化合物的过渡态，计算结果表明，两底物的催化机理和与活性部位的作用都相同。Caldwell 等又在底物的 P_1位用 Gly，Ala 和 Leu 进行点突变，计算 FE，结果 Ala 的过渡态最稳定，结合性最好。Hirono 等[94]用 FEP 方法计算了核糖核酸酶 T_1与其抑制剂 2′GMP 和 2′AMP 的结合性，计算结果表明，2′GMP 的结合性比 2′AMP 好，$\Delta G_{solv} \sim 29.26$ kJ/mol，$\Delta G_{bind} \sim 41.8$ kJ/mol。ΔG_{solv}计算表明，就疏水作用来说，A 的作用比 G 强，这使得蛋白质更易接近 A；Glu48 对 G 的选择性起主要作用，因 G 与之形成两个氢键。由计算得出结论，FEP 计算的前提是配体必须以相同的方式作用于受体的同一活性部位。

运用循环(2.6.7)，还可以研究酶的催化活性，只要将 X 当作底物，Y 当作酶，Y′当作残基定点突变后的酶。Rao 等[95]应用这一方法研究了胰蛋白酶骨架上两个 NH 组成的氧负离子孔的性质。他们将胰蛋白酶的 Asn155 用 Ala 定点突变，计算结果为 $\Delta\Delta G_{bind} = 0.42 \pm 3.34$ kJ/mol，$\Delta\Delta G_{cat} = 14.2 \pm 4.18$ kJ/mol，与实验值 1.67 ± 15.88 kJ/mol 非常符合。Hwang 等[96]用更加严格的量子力学方法优化过渡态，然后再用 FEP 计算 $\Delta\Delta G_{bind}$和 $\Delta\Delta G_{cat}$，得到相似的结论。Rao 等的计算结果表明，Thr221 与氧负离子过渡态作用非常重要，与 Ala 相比，Thr221 对过渡态的稳定作用提高 20 倍，这一结论已被实验证实[97]。

(2) 序列稳定性

可用下列循环研究核酸或蛋白质的序列稳定性[9]，

$$\begin{array}{ccccc} SS & \xrightarrow{\Delta G_1} & B & \xrightarrow{\Delta G_5} & Z \\ \downarrow{\scriptstyle \Delta G_3} & & \downarrow{\scriptstyle \Delta G_4} & & \downarrow{\scriptstyle \Delta G_7} \\ SS' & \xrightarrow{\Delta G_2} & B' & \xrightarrow{\Delta G_6} & Z' \end{array}, \qquad (2.6.62)$$

其中 Z 为 ZDNA，B 为 BDNA，SS 为单股 DNA。因为不知道核酸变化(B

$\rightarrow Z$)的途径，计算 ΔG_2和 ΔG_4非常困难，运用上述循环，即可计算 $SS\rightarrow B$、$B\rightarrow Z$ 的自由能变化，

$$\Delta\Delta G(SS \longrightarrow B) = \Delta G_2 - \Delta G_1 = \Delta G_4 - \Delta G_3, \qquad (2.6.63)$$

$$\Delta\Delta G(B \longrightarrow Z) = \Delta G_6 - \Delta G_5 = \Delta G_7 - \Delta G_4, \qquad (2.6.64)$$

从而知道 DNA 的稳定性。

用相似的办法可以研究蛋白质的序列稳定性，N 代表自然蛋白，D 为变性蛋白：

$$\begin{array}{ccc} N & \xrightarrow{\Delta G_1} & D \\ \downarrow{\scriptstyle \Delta G_3} & & \downarrow{\scriptstyle \Delta G_4} \\ N' & \xrightarrow{\Delta G_2} & D', \end{array} \qquad (2.6.65)$$

$$\Delta\Delta G(N \longrightarrow D) = \Delta G_2 - \Delta G_1 = \Delta G_4 - \Delta G_3。 \qquad (2.6.66)$$

Pearlman 等[98]用上述循环计算了因胞嘧啶被 5 -甲基胞嘧啶(5MEC)取代后，DNA 由 B 向 Z 转变的势差现象。他们分别计算了真空和溶剂中的自由能变化。先前的分子力学计算表明 5MEC 的势差可用核酸内相互作用来解释，并且 X 衍射晶体结构表明，由于疏水作用，5MEC 深埋在 DNA 的内部，Pearlman 的计算结果能完美解释这些现象，$\Delta\Delta G$ 计算结果与实验值一致。比较气相和溶液相的计算结果，$\Delta\Delta G$ 大约一半是分子内部相互作用的贡献，一半是自由能与溶剂效应的贡献。

许多人应用热力学循环和 FEP 方法研究了其他蛋白质的稳定性。Dang 等[99]用 FEP 和循环(2.6.65)计算了 T4Lyzozyme 酶 Thr157 突变为 Val 和 Ala 的自由能变化，计算结果重现了实验结果，Thr 突变为 Val 后，蛋白质稳定性降低，主要是由于 Thr 中甲基占了原有蛋白质 OH 的位置，增加了范德瓦耳斯排斥作用。

3. 量子化学计算方法应用

(1) 酶催化反应机理研究

碳酸酐酶(CA)是一种含锌金属酶，在人体或其他生物体中催化 CO_2 的可逆水合反应生成碳酸根离子。

$$CO_2 + H_2 \xrightarrow{CA} HCO_3^- + H^+。 \qquad (2.6.67)$$

CA 存在三种同工酶——CAI、CAII 和 CAIII，其中人体碳酸酐酶 II

(HCAII)催化活性最高，它的催化转换数为10^6/s。人体CA的三维结构已被测定[100]，它是由260个氨基酸残基组成的体积约为4.1×4.1×4.7 nm^3的椭球体。CA活性部位是一个圆锥体空穴，宽约为1.5 nm，深度约为1.6 nm，Zn^{2+}位于圆锥体的顶部，三个组氨酸His94、His96和His119与Zn^{2+}配位键合，溶剂水分子是Zn^{2+}的第四个配体(见图2.6.6)。活性部位被包含Thr199、Gln92和锌离子的平面分成亲水性区域和疏水性区域两部分。亲水性区域含有质子受体His64和部分有序的水分子，疏水性区域含有Val143、Trp209、Val121、Phe131和Leu141组成的CO_2结合"口袋"(见图2.6.6)。

HCAII存在于红细胞和人眼的睫状体中，在体液的分泌过程中起着重要作用。这一反应如失去平衡，会导致人眼中分泌的液体增多，眼压随之增高，这是青光眼产生的主要原因[101]。因此，CA催化CO_2水合反应机理的理论研究，可为设计高效的CA抑制剂、寻找治疗青光眼疾病的药物提供理论依据。

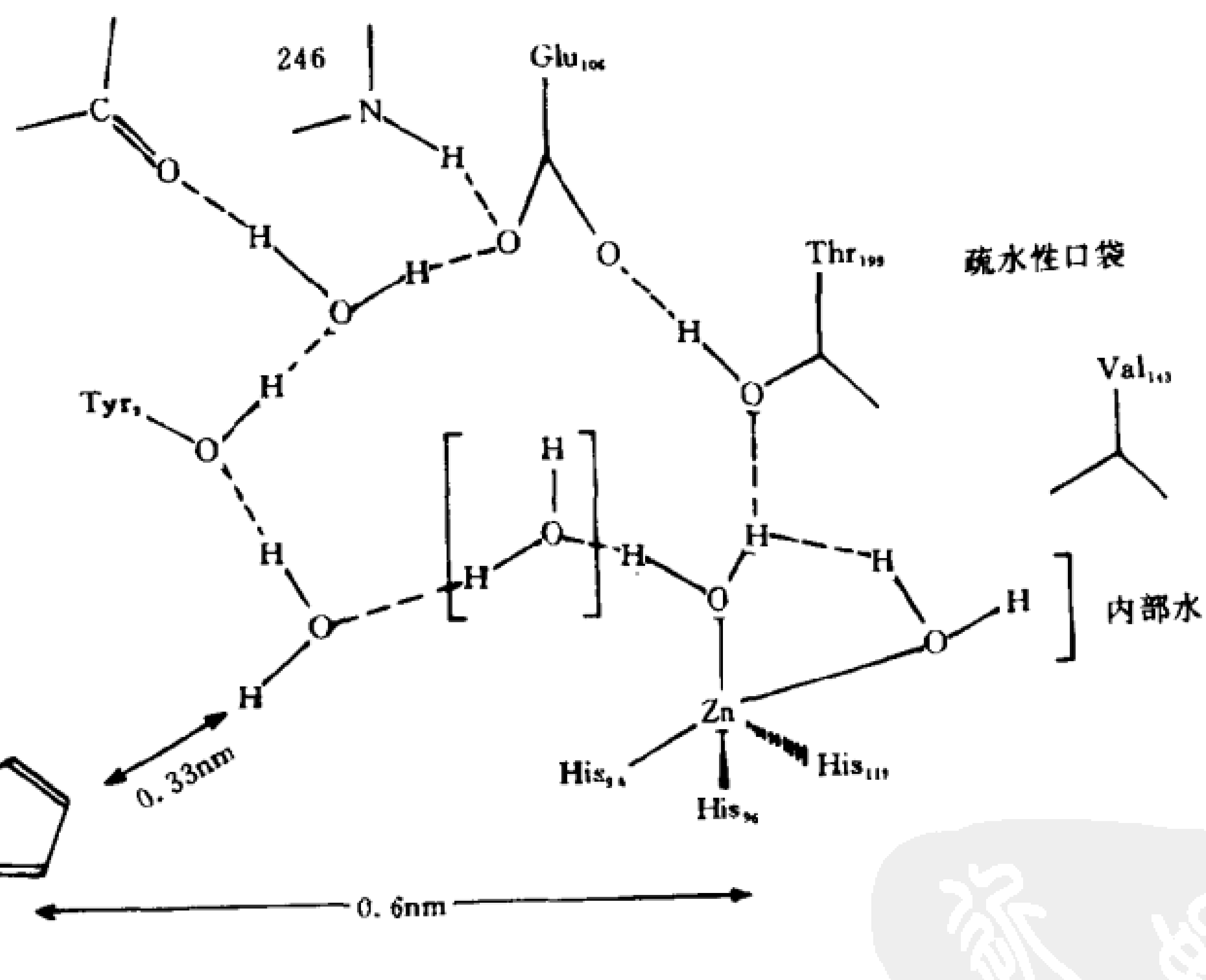


图2.6.6　人体CAII活性部位结构示意图

比较公认的反应机理是与Zn^{2+}离子键合的OH^-亲核进攻CO_2，其机理见图2.6.7。CA催化CO_2水合反应的机理分为四步，但对于第三步，即

$EZnHCO_3^-$ 的形成，到底是 Lipscomb 机理（图 2.6.7(B)）还是 Lindskog 机理（图 2.6.7(A)）存在着分歧。在 Lipscomb 机理中，$EZnOH^-$ 与 CO_2 亲核反应之后，在与羰基 C 相连的两个 O 原子之间有一质子转移，但 Lindskog 机理不存在上述质子转移，而是羰基 O 原子直接与 Zn^{2+} 键合。为了更清楚地了解 CA 催化 CO_2 水合反应的机理，许多化学家用量子化学方法对上述机理进行了理论研究[102]。

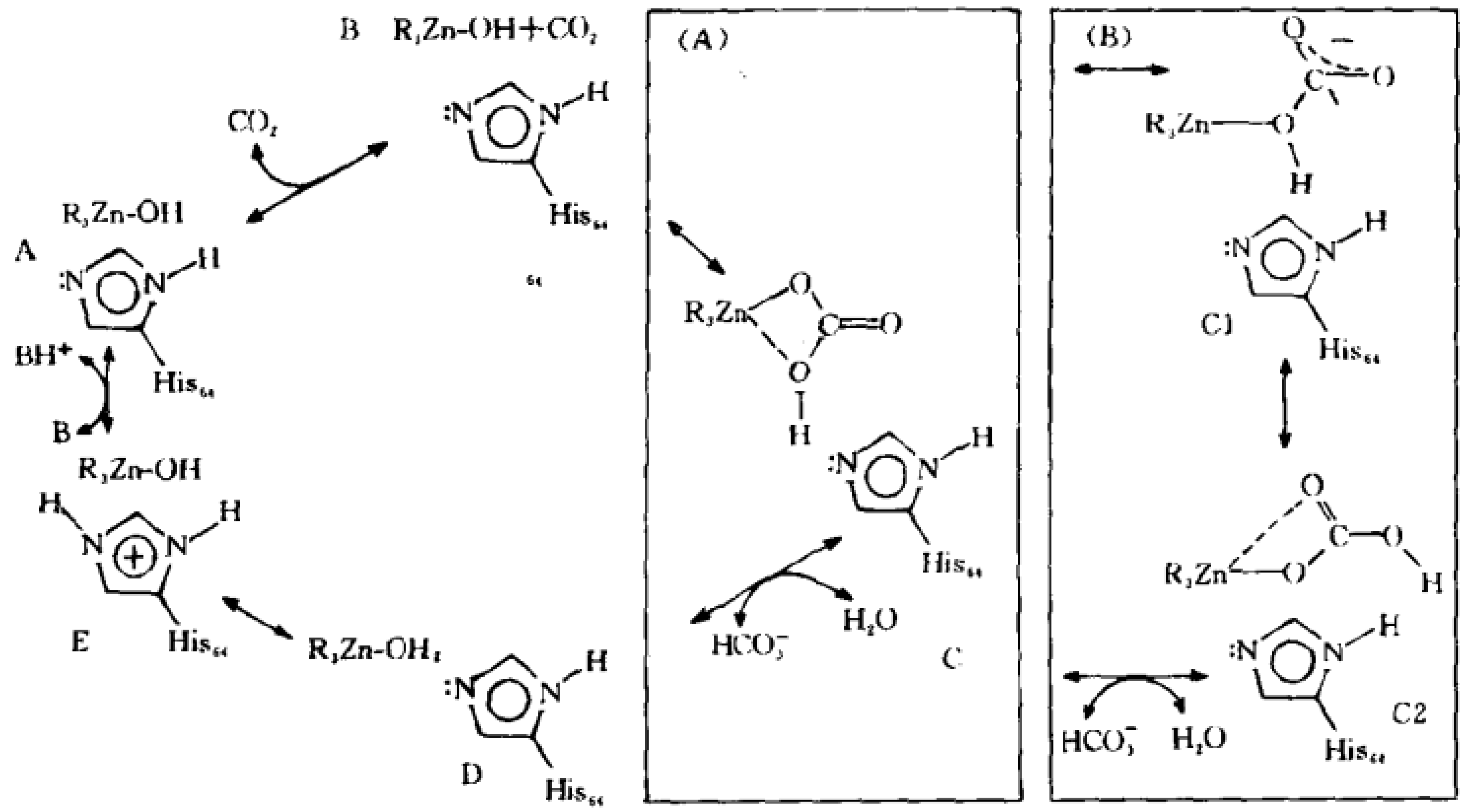


图 2.6.7 CA 催化 CO_2 水解机理示意图

Liang 和 Lipscomb[103~105]用 PRDDO 方法计算了 CA 催化 CO_2 水合反应的过程，其结果倾向于 Lipscomb 机理，并得到如下结论：$EZnOH^-$ 亲核进攻 CO_2 生成 $EZnHCO_3^-$ 的活化能为 0 kJ/mol；$EZnCHO_3^-$ 中两个 O 原子间质子转移活化能为 148.81 kJ/mol，如果引入一个或两个水分子作为质子传递的介质，则其活化能分别为 14.63 kJ/mol 和 5.852 kJ/mol；Glu106 - Thr199 链上的负电荷和水分子的电离作用使 H_2O 取代 HCO_3^- 的反应容易进行；Zn - H_2O 与 His64 间的质子转移降低 Zn - H_2O 的 pK_a，His64 或其他类似的质子受体的质子亲核性比 H_2O 强。

Merz 等[106]运用 AM1 方法，对 CA 催化 CO_2 水合反应的机理进行了详细的理论分析，得到了 Lindskog 机理的过渡态，活化能为 54.76 kJ/mol；但他们没有获得 Lipscomb 机理的过渡态，所以没对两种机理的优劣进行比较。Zheng 等[107]用 ab initio 方法研究了 CA 催化 CO_2 水合反应

的机理，并用 MD - FEP(molecular dynamics free energy perturbation)方法计算了环境对反应的影响，他们的计算结果倾向于 Lindsdog 机理。

Jacob 等[108]用 *ab initio* 方法，并用扩展基函数计算了 Lipscomb 机理和 Lindsdog 机理的活化能。结果表明，在 SCF 水平上，Lindsdog 机理的过渡态比 Lipscomb 机理的过渡态稳定，能量差为 209 kJ/mol，如考虑相关能，其能量差为 125.4 kJ/mol。

Sola 等[109]用 *ab initio* 方法研究了 Lipscomb 机理和 Lindskog 机理，发现气相计算结果倾向于 Lindskog 机理。他们还用自洽反应场(SCRF)方法计算了蛋白质残基和溶剂水分子对反应机理的影响，计算结果表明，在Lindskog机理中，H_2O 作为质子传递的介质，可以降低反应的活化能。因此，他们得出结论：如若考虑环境的影响，两种机理均有可能发生。

Krauss 和 Garmer[110]用 *ab initio* 方法研究了电离度对 CA 活性部位配体键合和 CA 催化CO_2水合反应的影响，酶的活性部位采用簇模型。由计算结果对 Zn—$OHCO_2$重排生成 HCO_3^- 提出了两种机理，一种是 CO_2的 O 原子进攻 Zn 直接重排生成 $ZnHCO_3^-$；另一种是在上述重排反应的过程中，在 HCO_3^- 和 H_2O 或 Thr199 之间进行质子转移。另外，他们还发现，活性部位残基负离子的存在，使得 Zn－OH^- 进攻底物羰基时放出的热增加，并有利于 HCO_3^- 的生成。他们的计算结果成功地解释了组氨酸质子 NMR 化学位移随 pH 值的变化以及 Michaelis - Menten 常数 K_m 不随 pH 值变化的实验现象。

许多化学家还用量子化学方法对其他酶的催化反应机理作了理论计算研究[108]。

(2) 多肽构象分析

丙氨酸二肽(Ala_2)长期以来一直是用来研究蛋白质或多肽平衡构象的一个模型化合物。虽然二肽构象分析的量子化学计算可以追溯到好几年以前，但只是到了 90 年代，才用 *ab initio* 方法进行了 Ala_2的构象分析。Head - Gordon 等[111]HF/3 - 21G 基组计算结果表明，Ala_2所对应的低能构象为 C_7^{eq}(1)、C_7^{ax}(3)、C_5(2)及能量略高的 α_R(5)(见图 2.6.8)。这一等值线图显示了 Ala_2二肽 Ramachandran 图的主要性质。许多人用 *ab initio* 对 Ala_2进行了研究，并计算了 $C_7^{eq}\to C_5$，$C_7^{eq}\to C_7^{ax}$，$C_7^{eq}\to\alpha_R$的转化过程，结果列于表 2.6.2。

表 2.6.2 用各种量子化学方法所计算的 Ala_2构象的相对能量(kJ/mol)

作者	文献	方法	C_7^{eq}	C_7^{ax}	C_5	α_R
		Hartree - Fock Results				
Gould and Kollman	112	HF/3-21G	0.0	11.7	4.60	
	112	HF/TVVP//6-31G * *	0.0	12.54	0.84	17.56
Bohm and Brode	113	HF/DZP	0.0	12.54	2.09	13.38
Head - Gordon, *et al.*	111	HF/6-31G *	0.0	10.87	0.84	
Frey, *et al.*	114	HF/6-31G * *			0.84	
		Estimates of Correlated Energies				
Weiner, *et al.*	115	HF/4-31G+DC	0.0	9.20	7.94	
Head - Gordon	111	MP2/6-31+G * //HF/6-31+G *	0.0	9.20	4.60	
Frey, *et al.*	114	MP2/6-311G * * //HF/6-311G * *	0.0		5.43	
Frey, *et al.*	114	MP2/6-311G * *	0.0		7.11	
Gould and Kollman	112	MP2/TZVP//HF/6-31G * *	0.0	8.78	6.27	16.72

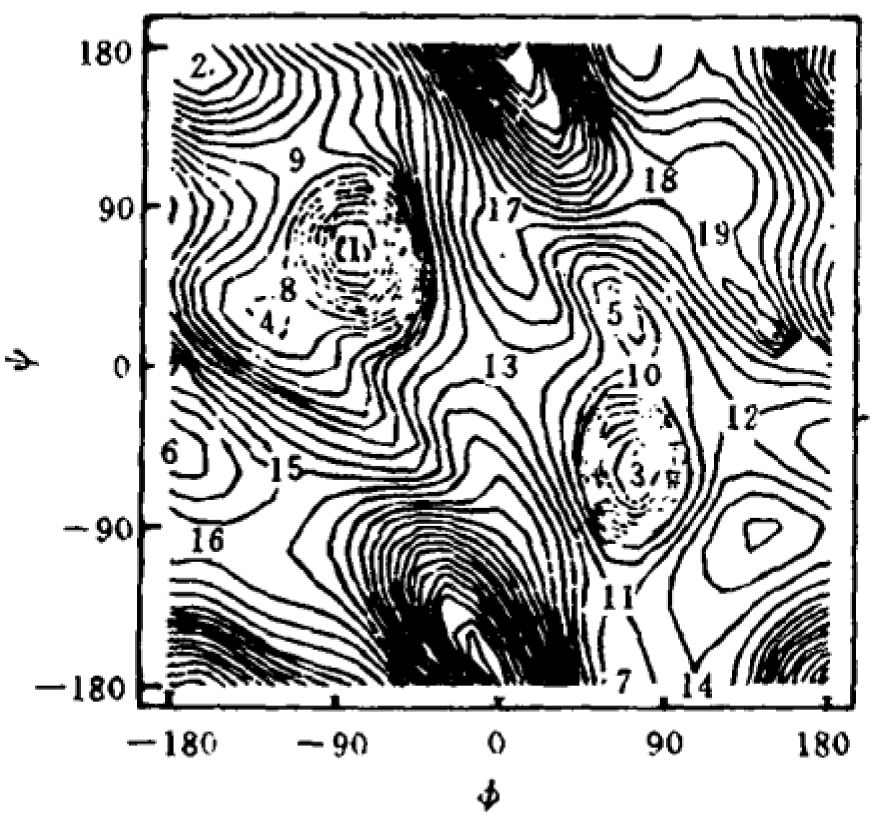


图 2.6.8 Ala_2构象能等值线图

($HF/3-21G$ 计算结果,$C_7^{eq}(1)$,$C_7^{ax}(3)$,$C_5(2)$,$\alpha_L(5)$)

4. 量子力学与分子力学相结合的 QM/MM 模型方法

Waszkowycz 等[117,118]应用 QM/MM 方法,研究了磷脂酶 A_2催化酯和酰胺水解的机理。他们应用磷脂酶 A_2 的晶体结构,将小分子酯和酰胺氧负离子与酶活性部位结合,向反应物或生成物方向改变 C—O 键长,然

后用 QM - 31 - 2G 方法进行构型优化,得到了水解反应的途径。如前所述,因 *ab initio* MO/MM 方法所能计算的 QM 较小,而半经验量化的计算精度可与小基组从头算法相比,Field 等半经验 MO 方法 AM1 加入到 MM/MD 程序 CHARMM 中,Brooks 等[119]用此方法研究了磷酸丙糖异构化酶(TIM)的催化反应,计算结果重现了 Knowles 等[120]提出的反应机理,理论计算能较好地解释实验现象。

5. EVB 模型

EVB 模型也是一种 QM/MM 模型,是将量子力学中的价键理论与分子力学(MM)或分子动力学(MD)相结合的计算方法,与 MO/MM 计算方法相比,这一方法可将计算结果与实验数据拟合,调节参数,不但可以同气相实验数据拟合,而且也可与溶液中的实验参数拟合。这是 EVB 理论的优点。现 EVB 理论被广泛地应用于生物大分子,特别是溶液中的化学反应和酶催化机理的研究工作[3,45]。

在量子力学计算方法的应用中,我们已对 CA 催化机理以及 QM 研究作了描述,这里对 Warshel 等[3,46]用 EVB 方法对 CA 催化机理的研究作一介绍。

Warshel 等[121]首先用 SCAAS 模型在 SPC 水环境中优化 EVB 计算 $H_{12}(r)$的参数,质子转移 VB 波函数为

$$\Psi_1 = (HOH)(OH_2), \quad \Psi_2 = (HO^-)(HOH_2^+)。 \tag{2.6.68}$$

通过计算得到 EVB 参数,然后用上述参数计算 HCAII 以 Zn^{2+} 为中心 1.4 纳米范围内残基环境中的 H_2O 间质子转换自由能变化。在水溶液中活化能和反应自由能计算值均与实验值相符合($\Delta G^{\neq}_{obs}=99.48$ kJ/mol, $\Delta G^{0}_{obs}=89.45$ kJ/mol)。在酶环境中,质子转移的活化能为 45.14 kJ/mol,反应自由能为 41.8 kJ/mol,也与实验值一致。他们还用统计力学方法计算了质子转移的反应速率常数 k,$k\approx 7\times 10^4\ s^{-1}$,上述速率常数的计算没有考虑量子效应。但这一计算值与实验值相符($6.8\times 10^4\ s^{-1}$)。

在计算 CO_2在酶催化下的水化反应时,Warshel 等[121~123]首先用在水溶液中 CO_2水化反应的实验值调节 EVB 参数,其 VB 波函数为

$$\Psi_1 = (HO^-)(CO_2), \quad \Psi_2 = (HOCO_2)。 \tag{2.6.69}$$

他们得到了 OH^-亲核进攻 CO_2的能量曲线,OH^- 与 CO_2在水溶液中反应的活化能为 49.74 kJ/mol,反应自由能为 −43.89 kJ/mol。EVB 计算值

与实验值一致。用水溶液中的 EVB 参数，他们计算了酶催化下的 CO_2 水合反应的自由能，计算结果为亲核进攻的活化能，在酶环境中比在溶液中降低了 20.9～25.08 kJ/mol，并且这步反应所放出的热量比水溶液中的要少。$\Delta G^{\neq}$ 和 ΔG_0 的计算值分别为 26.33 kJ/mol 和 −20.06 kJ/mol。另外他们还比较了过渡态 EVB 计算结果和量子化学计算模型化合物的过渡态结构，两者的几何构型非常相似。EVB 理论还被用于其他酶的催化反应中，如：溶解酶、胰蛋白酶和其他丝氨酸蛋白酶[46]。

6. SIBFA

SIBFA 被广泛应用于 DNA 与药物的结合性研究中[124]。Gresh 等[125]用 SIBFA 研究了抗癌药物 7－H 吡啶并咔唑 D－Asn 衍生物与 DNA 的作用方式，第一种是吡啶并咔唑插在 $d(CpG)_2$ 和 d(CpA)·d(TpG)序列中，第二种是吡啶并咔唑环插在 $d(GpC)_2$ 和 d(GpT)·d(ApC)序列中。对于第一种模式，最好的结合序列为 d(ACG)·d(CGT)和 d(ACA)·d(TGT)。在这一复合物中，与吡啶并咔唑相连的肽支链几乎与 DNA 双螺旋的轴平行，D－Asn 与 DNA 形成二重氢键(D－Asn 的 C=O 和 NH，DNA 腺嘌呤碱基的 HN_6 和 N_7)。对于第二种插入方式，吡啶并咔唑环的凹面向 DNA 螺旋链，二个烷基季铵盐与 3′鸟嘌呤上的 N_7 和 O_6 形成氢键。肽链与含有 A、G 或 C 的胍基结合，这一结合方式从能量上来看与最好的(嘧啶-嘌呤)三聚体相近。根据计算结果，他们设计并合成了一些类似物，实验结果与计算一致。同时 Gresh 等研究了水溶性纺锤素三正离子衍生物与 B－DNA $d(CGCGAATTCGCG)_2$ 和 $d(GCGCAATTGCGC)_2$ 的相互作用，并根据计算结果成功地合成了其中一些化合物。Gresh 等根据 SIBFA 计算结果设计了 Mitoxantone 双四肽衍生物。在反转录病毒(retroviruses)和 ptoyoonvohrnrd 过程中，六聚核苷酸 $d(GGCGCC)_2$ 以周期性的方式转录激活序列。

最近 Gresh 对 SIBFA 进行了改进[126]，使原来只能进行 DNA -配体相互作用计算的 SIBFA 能计算含锌蛋白质与配体的相互作用[126]。Gresh 用改进的 SIBFA，计算了 Zn^{2+} 与金属酶中常见的配体——水、羟基、羟甲基、甲硫醇、脒唑等组成的配合物[126]，并与超分子从头计算结果[127]比较，以验证 SIBFA 计算结果的可靠性。然后计算嗜热菌蛋白酶与其抑制剂的相互作用，得到了较好的结果。

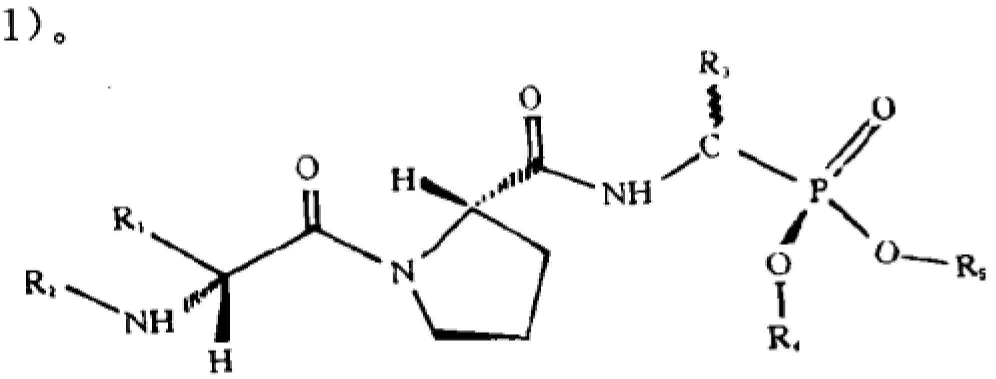
2.6.4 我们关于配体-受体相互作用和药物设计的研究工作

近年来，我们也开展了生物大分子体系及其相互作用的理论计算和基于受体三维结构的合理药物设计工作，现将部分工作作一简单介绍。

1. 凝血酶抑制剂与人体α凝血酶相互作用的理论研究

凝血酶是一种胰蛋白丝氨酸蛋白酶，在凝血和血栓形成中起着重要作用[16]。在凝血过程中，凝血酶催化血纤维蛋白原水解成血纤维蛋白肽A和B，血纤维蛋白肽A、B再聚合形成血栓。体内凝血酶生理功能异常是心脑血管的主要病因之一。凝血酶抑制剂可以治疗脑血栓、心肌梗塞和动脉粥样硬化等疾病，有些凝血酶抑制剂还可以作为抗炎药物和抗癌药物。但现有的凝血酶抑制剂都是多肽类化合物，结构较为复杂，如cyclitheonmides A和B是一个19元环的多肽，水蛭素是65肽，它们均难以合成，且存在作用时间短、选择性差等缺点。寻找新的高活性和长效凝血酶抑制剂仍是一个尚待解决的问题[128]。凝血酶的三维结构已经阐明[129]，通过研究现有抑制剂与凝血酶的相互结合与作用，特别是寻找其活性构象和药效基团，有可能设计出结构简化、易于合成的新的小分子抑制剂。

我们用自编程序BCSPL[130,131]，根据人体α凝血酶活性部位的结构搜寻了磷酰化肽类凝血酶抑制剂D-Aa-Pro-Aa^P(图2.6.9)的活性构象(图2.6.10)。然后用分子力学方法优化了抑制剂-凝血酶复合物的结构(图2.6.11)。



1. R_1=$PhCH_2$,R_2=H,	R_3=*n*-pentyl(R),	R_4=R_5=Ph
2. R_1=Ph_2CH,R_2=H,	R_3=*n*-pentyl(R),	R_4=R_5=Ph
3. R_1=$PhCH_2$,R_2=H,	R_3=3-methoxypropyl(R),	R_4=R_5=Ph
4. R_1=Ph_2CH,R_2=H,	R_3=3-methoxypropyl(R),	R_4=R_5=Ph
5. R_1=$PhCH_2$,R_2=PhCOO,	R_3=*n*-pentyl(R),	R_4=R_5=Ph
6. R_1=$PhCH_2$,R_2=H,	R_3=*n*-pentyl(R),	R_4=R_5=H
7. R_1=Ph_2CH,R_2=H,	R_3=*n*-pentyl(R),	R_4=R_5=H
8. R_1=Ph_2CH,R_2=H,	R_3=*n*-pentyl(S),	R_4=R_5=H

图2.6.9 磷酰化肽类凝血酶抑制剂

D-Aa-Pro-Aa^{P}结构示意图

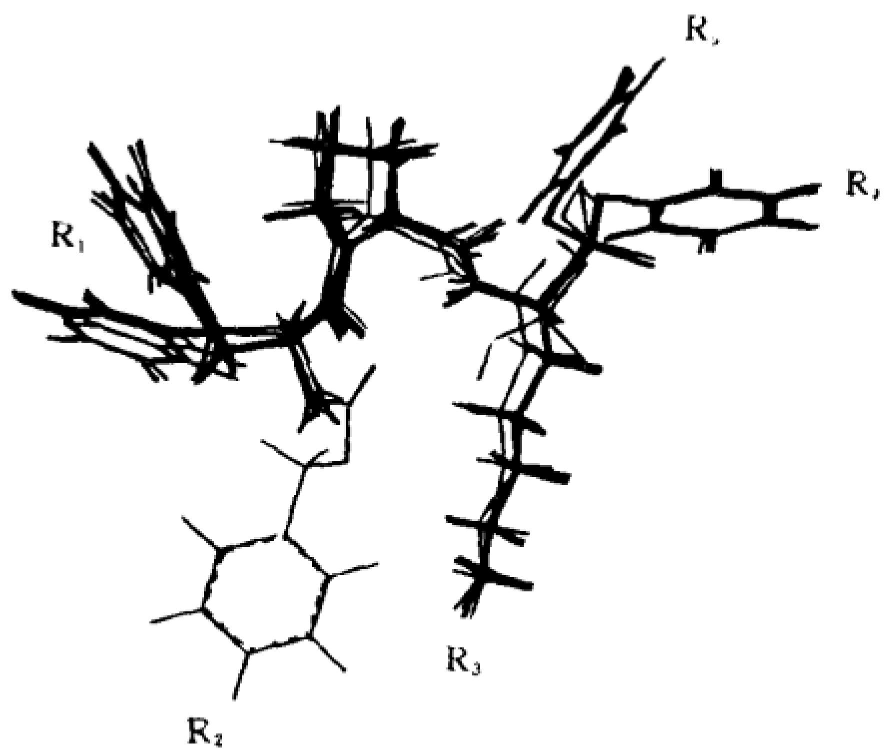


图 2.6.10 磷酰化肽类凝血酶抑制剂的活性构象及其在凝血酶活性部位的叠合

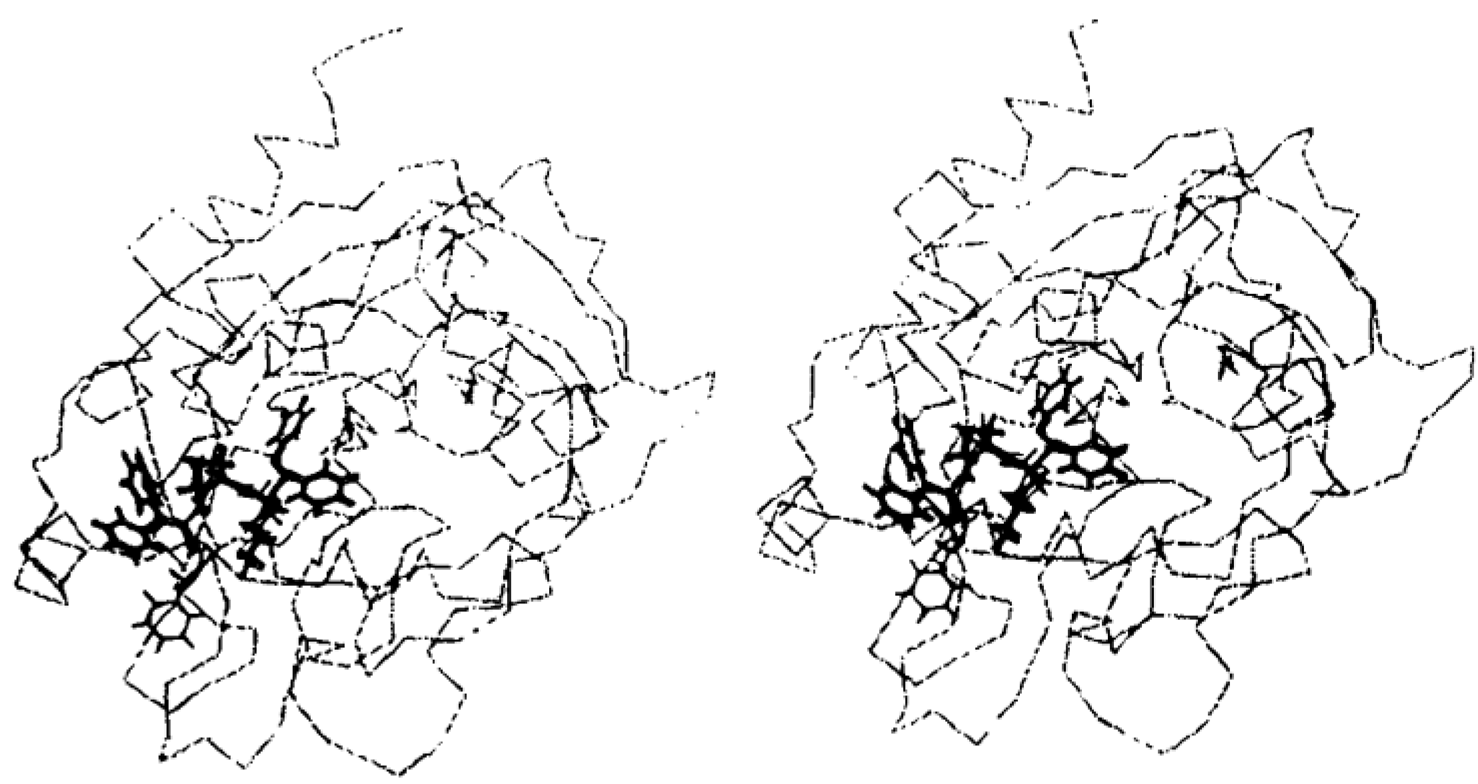


图 2.6.11 磷酰化肽类凝血酶抑制剂-凝血酶复合物分子力学优化结构(立体图)

由计算结果,我们发现,活性 $-\log IC_{50}$ 与抑制剂-凝血酶的相互作用能 E^{b}_{total} 有较好的相关性(图 2.6.12),其回归方程为:

$$-\log IC_{50} = 1.417 - 0.035E^{b}_{total},$$
$$r^2 = 0.786,\quad F = 14.694,\quad s = 0.746。\qquad (2.6.70)$$

由抑制剂-凝血酶复合物的结构,我们得到了磷酰化肽类凝血酶抑制剂与凝血酶氢键相互作用的模式,并根据这一模式推测了磷酰化肽类凝

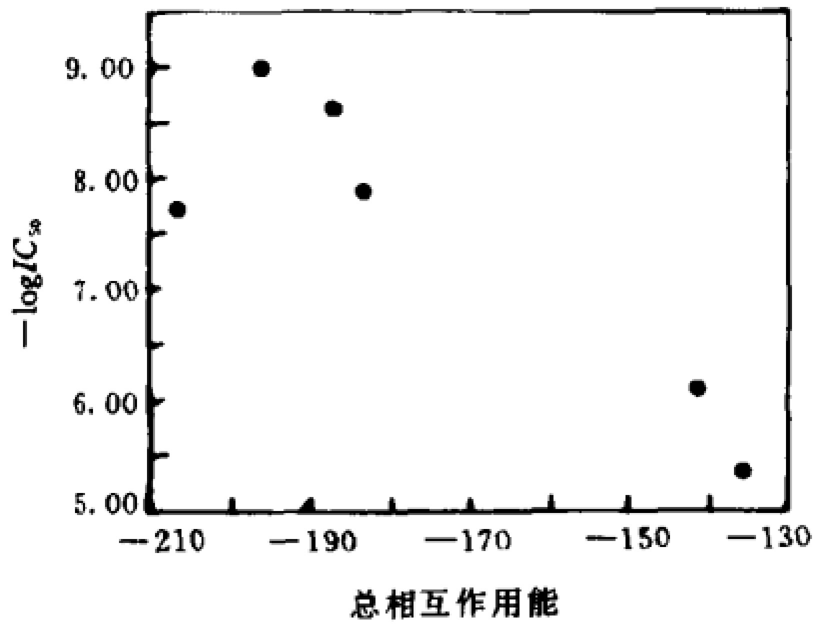


图 2.6.12　活性与相互作用能的相关性

血酶抑制剂的抑制机理。

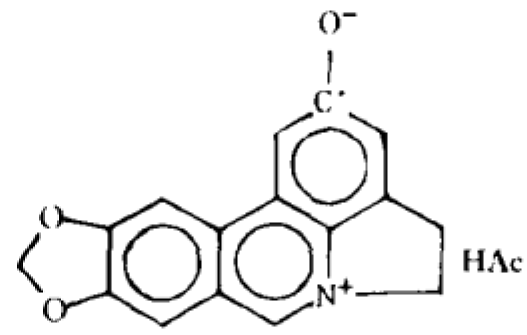
根据磷酰化肽类凝血酶抑制剂的活性构象及其在凝血酶活性部位的迭合方式，我们用比较分子力场分析(CoMFA)方法，对这些抑制剂进行了三维定量构效关系(3D-QSAR)分析，得到了较好的 3D-QSAR 模型，其力场分布与凝血酶活性部位的结构基本一致。用类似方法，我们还对二肽、三肽类凝血酶抑制剂 R—Pro—Arg—H 与凝血酶的作用机理进行理论研究，并进行了 3D-QSAR 分析[130]。

在上述理论计算和 3D-QSAR 的基础上，我们进行了新型凝血酶抑制剂的设计工作，所设计的分子合成工作正在进行之中。

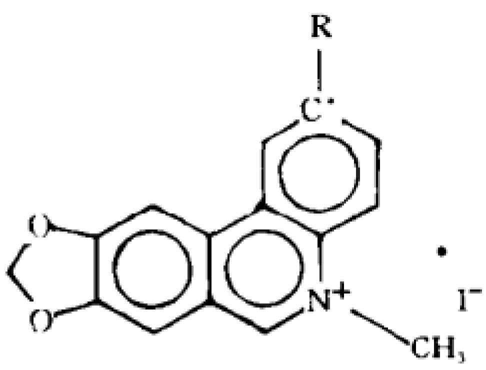
2. *抗癌药物-核酸相互作用的研究*

许多抗癌药物是 DNA 嵌入剂，研究这些嵌入剂与 DNA 的相互作用，能为新型抗癌药物的设计提供一定的线索，DNA 或寡聚核苷酸经常作为抗癌药物设计的"靶标"。近年来，我们在抗癌药物-DNA 相互作用理论计算的基础上[132,133]，用量子化学和分子力学方法以及多维核磁共振技术，对石蒜内铵及其类似物与 DNA 的相互作用进行了较为系统的理论计算和分子设计研究。石蒜内铵(AT-1840，图 2.6.13)是我所合成的抗肿瘤药物，实验发现，石蒜内铵是一种 DNA 嵌入剂。

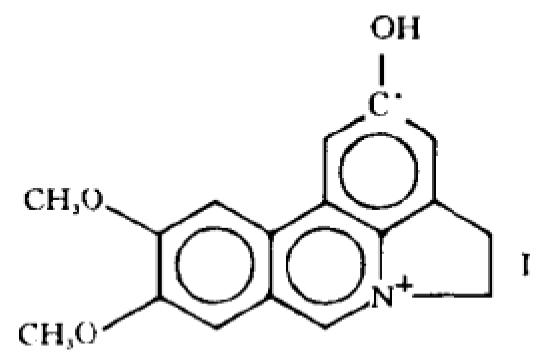
我们选择了六组双螺旋四聚核苷酸 d(GCGC)$_2$、d(CGCG)$_2$、d(ATAT)$_2$、d(TATA)$_2$、d(GTGT)$_2$和 d(ACAC)d(CCGG)作为石蒜内铵作用的受体，用 SIBFA 优化了石蒜内铵-寡聚核苷酸的结构，并计算了它们之间的相互作用能[134]。计算结果表明：



AT－1840



R
s121：$-CH_2OH$
s133：$-COCH_3$
s138：$-NO_2$
s140：$-Cl$



s74

图 2.6.13　石蒜内铵及其类似物的结构示意图

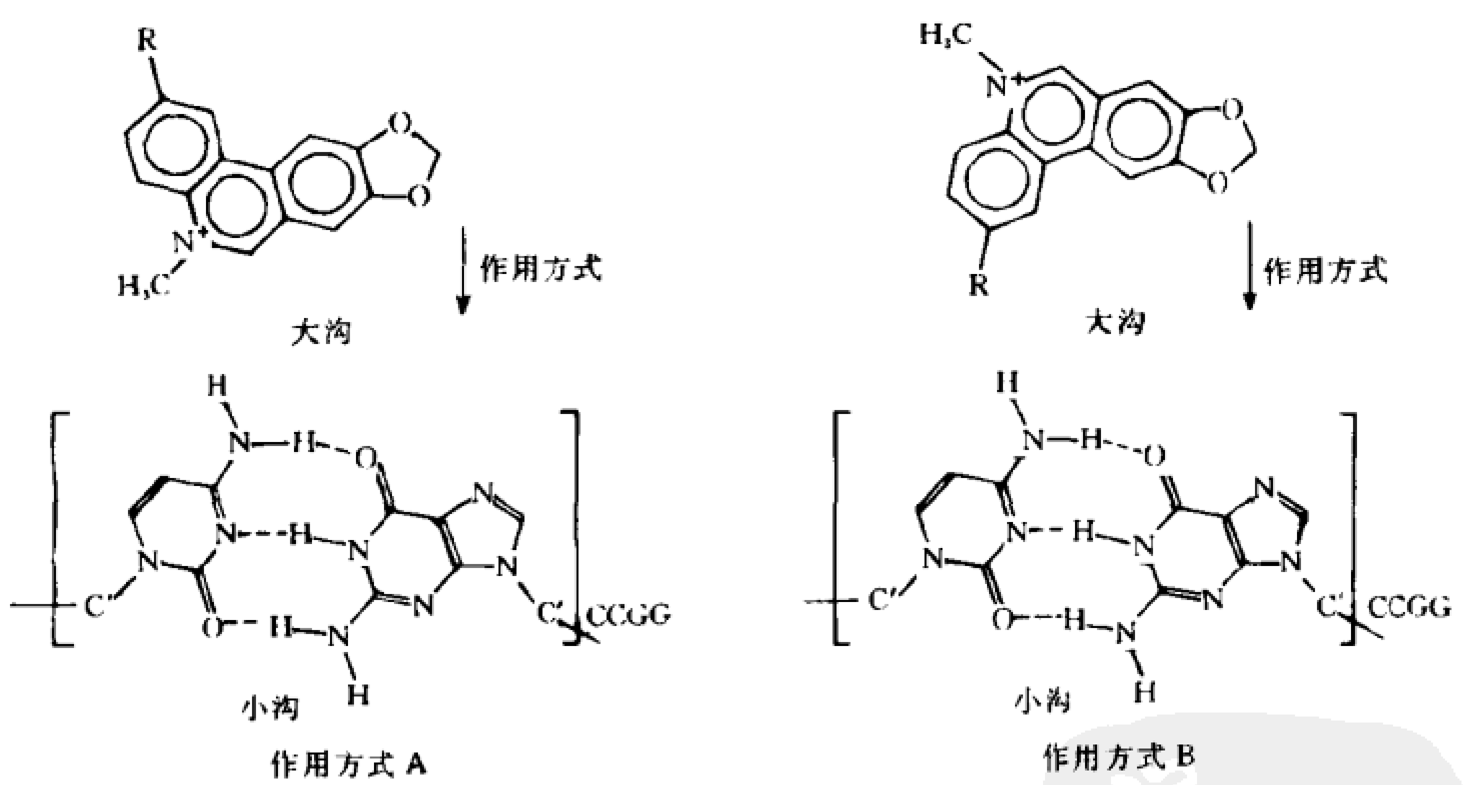


图 2.6.14　石蒜内铵类似物与寡聚核苷酸的作用方式示意图

① AT－1840 可能以 A、B 两种方式与 DNA 作用，且以与上下两对碱基平行的方式插入两对碱基之间(见图 2.6.14)；

② AT－1840 与嘧啶(3′－5′)-嘌呤的结合能力要比与嘌呤(3′－5′)-

嘧啶的结合能力强；

③ 在上述两种方式中，化合物对 d(CCGG)$_2$的嵌插作用均较其他几种碱基序列强。

在上述计算的基础上，我们又对不同石蒜内铵类似物(见图 2.6.13)与寡聚核苷酸的相互作用进行了理论计算[135]，结果表明：

① 这些化合物与 d(CCGG)$_2$的结合能与其抗肿瘤活性显著相关，两者基本平行；

② 各类似物都以“大沟方式”与 d(CCGG)$_2$的相互作用较为有利；

③ 次甲二氧基上的氢原子与核苷酸上的磷酸氧原子形成氢键，因此，如果该五元环受到破坏，将降低化合物与核酸的结合能；

④ 随着 C* 位上的侧链基团吸电子作用的加强，化合物与核酸的作用能也随之增加。

根据这些结论，我们设计了两种新的 AT－1840 类似物，它们与核苷酸的结合能明显强于现有的分子，因此预见它们将具有更强的抗癌活性。

3. *石蒜内铵-DNA 相互作用的核磁共振研究*

对药物的靶分子复合物进行 NMR 研究，可获得药物分子和靶分子在溶液中的构象，为药物设计提供如下重要信息：通过复合物构象了解药物各基团与靶分子结合时的空间位置；鉴别配基与靶分子相互作用的部位，区别作用必需基团；了解药物与靶分子作用的新区域。

根据药理研究和我们的计算结果，我们设计了一个新的寡聚核苷酸片段(CCGTACGG)$_2$，并采用液相合成法合成。NMR 谱在 Bruker AM-400 谱仪上测定，取 HDO 溶剂峰 $\delta=4.85\times10^{-6}$为图谱内标，溶剂峰用预饱和法抑制，FT 变换前经位移正弦窗函数处理，变换后再对称化。用计算机辅助解谱技术，我们从 NMR 测定的 NOE 结果，得到了石蒜内铵-(CCGTACGG)$_2$复合物氢原子的距离限制，用距离几何方法得到了复合物的初始结构，再用分子力学和分子动力学对初始结构精确化，得到了复合物的三维结构(图 2.6.15)[136]。实验结果表明：

① 石蒜内铵与 d(CCGTACGG)$_2$组成 2∶1 复合物；

② d(CCGTACGG)$_2$为 B－DNA；

③ 石蒜内铵的嵌入位点为 CpC 和 GpG 中的 CG 之间，这和我们先前的理论计算预测一致；

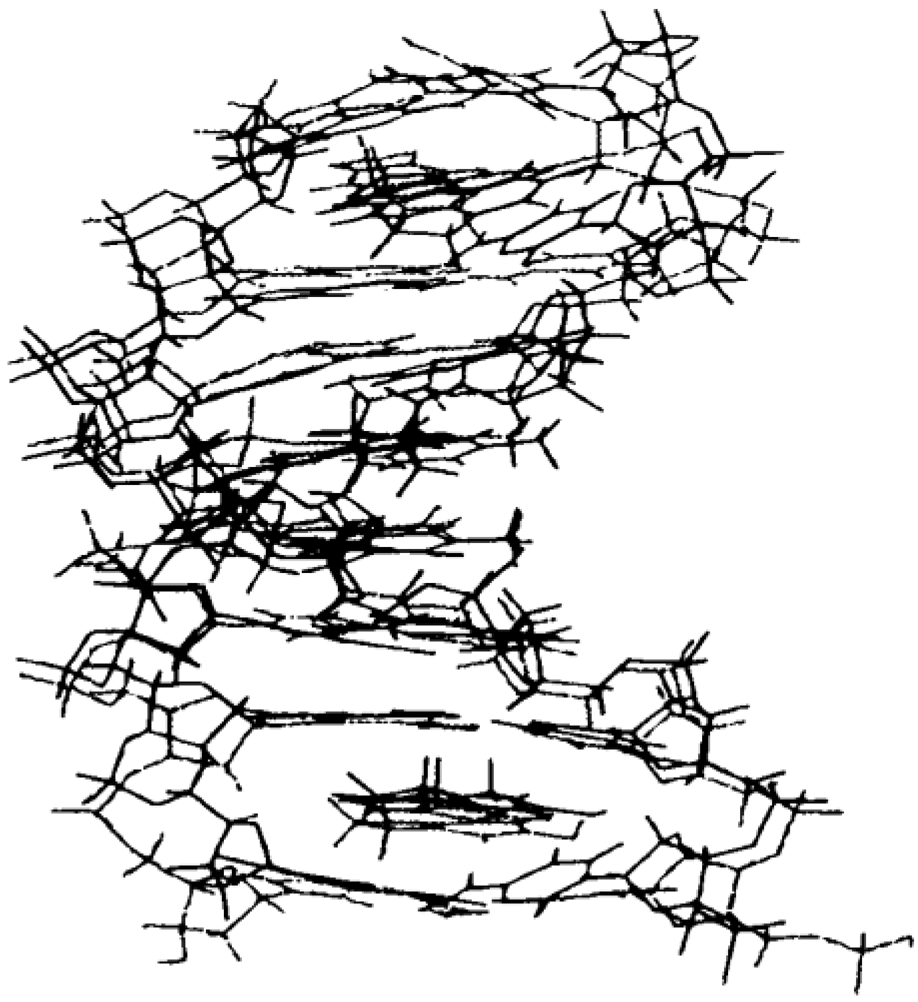


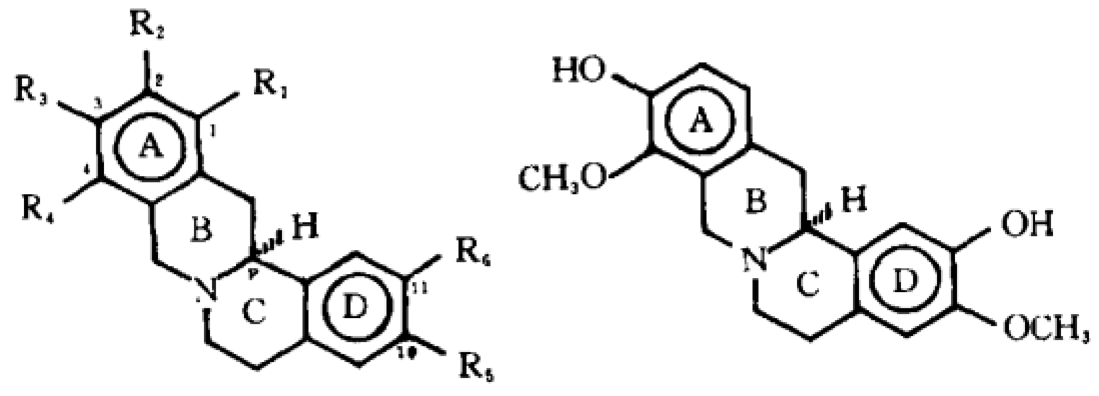
图 2.6.15 石蒜内铵和核酸复合物的溶液结构

④ 维持复合物结构稳定的因素中，季铵氮的正电荷是一个因素，但更重要的因素为药物和 DNA 间的范德华相互作用和疏水相互作用。

根据以上计算和实验结果以及石蒜内铵-$(CCGTACGG)_2$的溶液构象，我们设计了一些新的石蒜内铵类似物，并已完成部分合成工作，现药理测试工作正在进行之中。

4. 基于多巴胺受体三维结构模型的药物作用机理研究

多巴胺(DA)受体是治疗帕金森氏症和精神分裂症药物的作用靶标。采用基因克隆技术已识别至少五种多巴胺受体亚型，分属于D_1类和D_2类两个亚家族。它们都属于G蛋白偶联受体超家族，由一条单链跨膜7次而成，配体结合部位被认为在7个跨膜α螺旋束之间。四氢原小檗碱类似物(THPB，图 2.6.16)是我所从天然产物中发现的一类新型多巴胺受体阻滞剂[138]，它们对脑内D_1、D_2类受体均有亲和力，但优先作用于D_1类受体。为了设计出新的作用更强的药物，我们采用分子模拟技术，首先建立了多巴胺D_1、D_2类受体的三维结构模型，然后选择药理作用很强的左旋千斤藤立定(I-SPD)作为 THPB 的先导化合物，对其作用机理及其对各亚型的选择性进行了理论研究[139,140]。

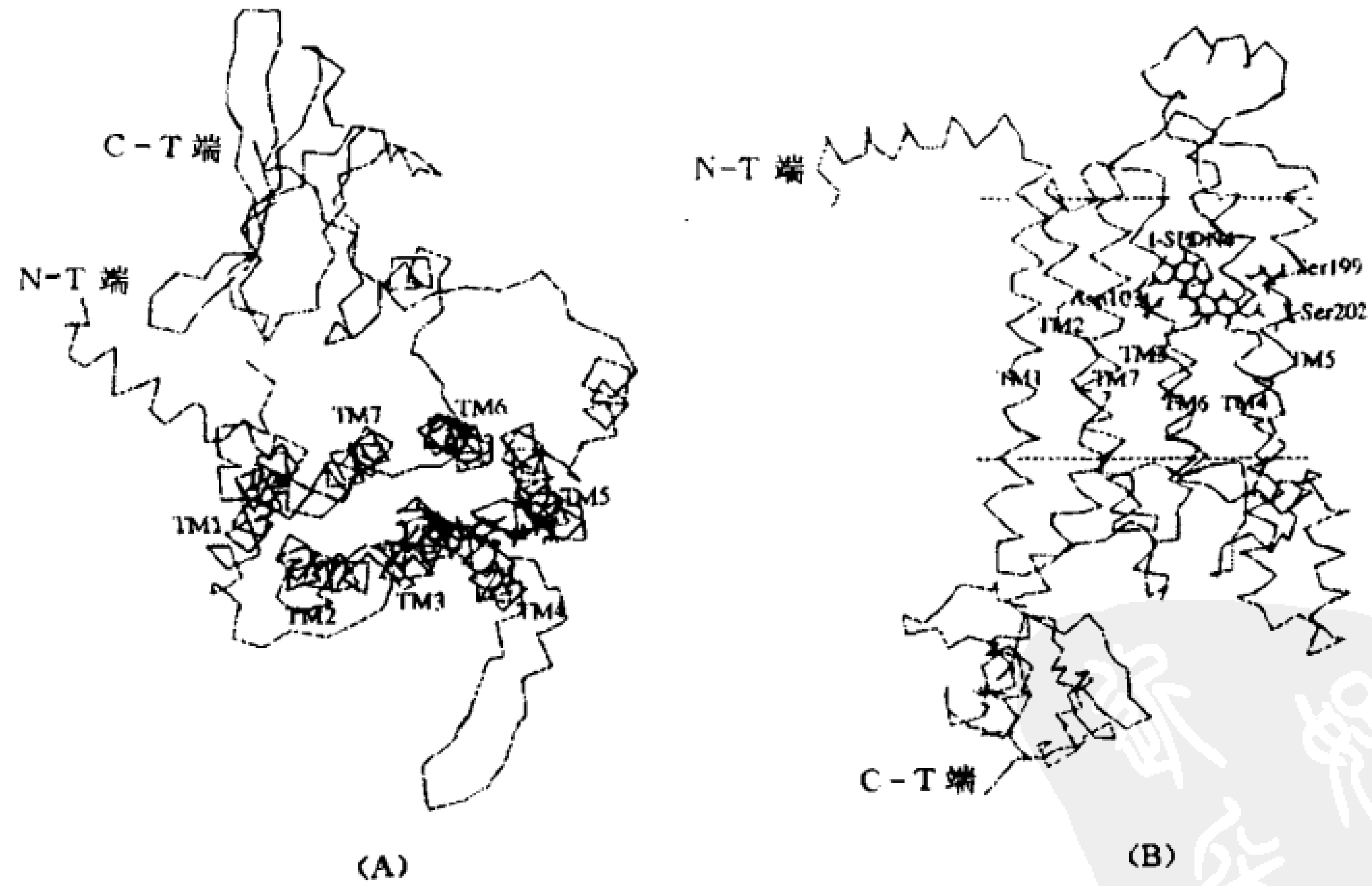


(A)　　(B)

图 2.6.16　(A)THPB 结构；(B)I－SPD 结构

(1) 相互作用的分子模拟研究

多巴胺受体的三维结构至今尚未阐明，因此首先要用分子模拟技术建立其结构模型。然而现有的蛋白质结构数据库(PDB)中缺乏与 DA 受体同源的蛋白结构，这给结构模建带来了困难。我们通过序列分析，首先以含七个 α 螺旋的细菌视紫红质的晶体结构[141]为模板，分别建立起人体 D_1、D_2类受体保守区的结构。以天然配体 DA 对接到其活性部位进行相互作用模拟，得到的结果与定点突变实验一致[142]。这证明保守区的结构具有可靠性，可用于 THPB 的作用机理研究。接着在 PDB 库中进行环区搜寻，得到了可变区的结构。可变区的结构虽尚有待证实，但有助于我们对



(A)　　(B)

图 2.6.17　质子化 I－SPD 与 D_1类受体所形成的复合物示意图

受体只显示 α 碳链。(A)从膜外垂直往内看；(B)从膜侧面观看。

受体整体结构的理解。使用分子力学进行结构优化以后，将 I－SPD 分别对接到 D_1、D_2类受体的活性部位，经过手工调整及结构优化，得到了 I－SPD 与受体的相互作用模型。由于配体在体液环境中易质子化，因此也建立了质子化 I－SPD 与受体的相互作用模型(图 2.6.17，2.6.18，表 2.6.3)。

表 2.6.3 配体-受体相互作用能(kJ/mol)

	D_1类受体			D_2类受体		
	总能量	立体能	静电能	总能量	立体能	静电能
I－SPD	－507.907	－501.533	－6.374	－374.277	－360.575	－13.702
质子化 I－SPD	－641.943	－490.280	－151.663	－510.695	－389.099	－124.953

我们从上述相互作用模型中发现，I－SPD 及 THPB 中质子化 N 与受体 TM3 中 Asp 残基形成静电与氢键作用，D 环上二取代基分别与受体 TM5 上两个 Ser 残基形成氢键作用，芳香环与周围几个芳香残基形成 $\pi-\pi$ 相互作用，所有这些相互作用使得 I－SPD 及 THPB 对 DA 受体具有较高的亲和力，从而具有较强的药效。从表 2.6.3 中可以看出，D_1与 D_2类受体的主要差别表现在立体作用能上，D_1类受体中围绕配体的芳香环较多，因而与 THPB 类化合物作用时具有更低的立体作用能，从而两者之间具有更高的亲和力。

(2) 相互作用的量子化学研究

为了深入研究其作用机理，我们对配体及受体分别进行了量子化学计算。使用半经验量子化学 AM1 法计算 30 个 THPB 左旋体及其质子化结构，发现所有 THPB 质子化前后的 HOMO 轨道均位于 D 环，LUMO 轨道质子化前位于 A 环、质子化后位于 N 原子并有小部分位于 A 环上，且 HOMO 轨道能与活性呈正比例趋势。在上述相互作用模型基础上，抽提出 D_1、D_2类受体的活性部位(图 2.6.18 所示)，使用 PM3 法进行半经验计算。结果表明，HOMO 轨道位于 D_1类的 Asp103(D_2类的 Asp114)上，LUMO 轨道位于 D_1类的 Ser202 及 Phe203(D_2类的 Ser197 及 Phe198)上，受体配体的前线轨道正好匹配。由前线轨道理论可知，配体的 D 环电子转移到受体芳香环上形成 $\pi-\pi$ 相互作用，该作用对分子识别起重要作用，配体的 LUMO 与受体的 HOMO 间产生强的静电与氢键作用。同时对同一配体而言，HOMO 能量一定，D_1类具有更低的 LUMO

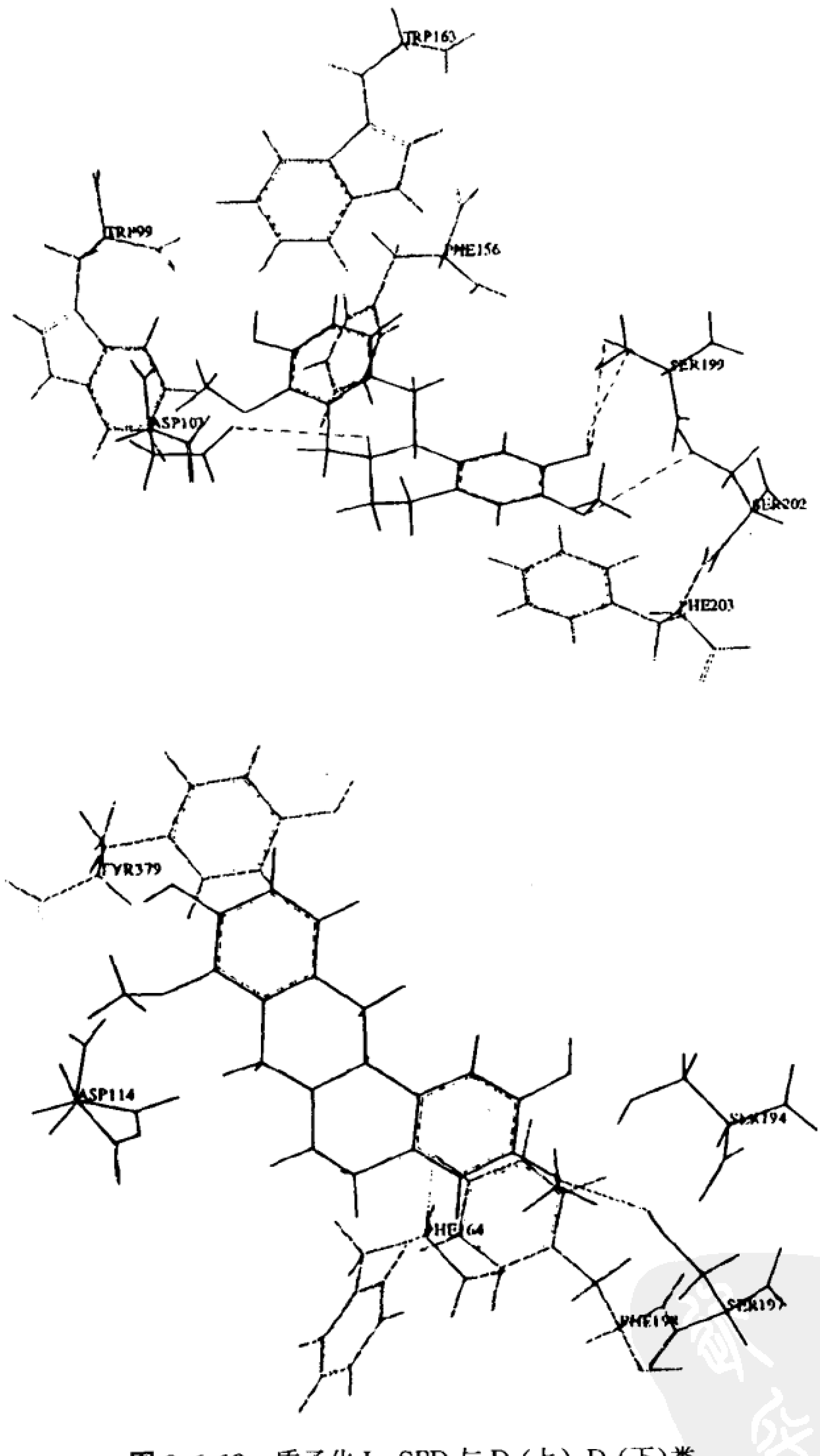


图 2.6.18 质子化 I-SPD 与 D_1(上)、D_2(下)类受体活性部位相互作用模式

能量(较 D_2类低 0.399 7 eV 即 38.567 3 kJ/mol),因而 D_1类更易接受电子,这可能是 THPB 优先作用于 D_1类的原因。

基于上述理论研究,我们提出如下作用机理:THPBs 与 DA 受体的相互作用可分为两步进行,首先进行分子识别,即由于静电取向作用使得配体与受体的前线轨道匹配;然后发生电荷迁移及其他相互作用,即 D 环与芳香残基的 π-π 相互作用,质子化 N 与 Asp 残基之间的静电与氢键作用,D 环上二个取代基与二个 Ser 残基之间的氢键作用。由于 D_1类受体更易接受电子,因而其 π-π 相互作用更强,使得 THPBs 与之优先作用。

2.6.5 展望

1. *理论计算方法*

从上面的讨论可以看出,每一种方法都有其一定的应用范围,并且要研究某一生物体系,仅用一种方法往往是不够的,而必须联合使用几种方法。理论计算已成为现代生命科学研究中不可缺少的一部分。

(1) 大体系量子化学计算

过去由于计算条件和计算方法的限制,量子化学所能计算的体系较小,现在由于计算条件的改善,已可计算较大体系的分子,其中酶催化机理的研究是一个热点。酶催化反应的 QM 研究关键要选好模型化合物,在选模型化合物时要尽量接近天然酶活性中心的真实结构,并加入环境的影响,向超分子量子化学计算发展。

密度泛函理论可以计算的体系比 ab initio 大,目前在材料分子的计算中应用较广,但在生物学中的应用还刚刚开始;同样 ab initio 赝势法计算的体系更大,目前也只适合于对称性高的材料分子。由于生物分子的柔性通常很大,ab initio 赝势法在生物学中的应用还不广,但这一方法今后可能会更多地应用于生物分子的计算。

(2) QM/MM 结合的方法

QM/MM 相结合的方法是目前理论计算的又一热点。在生物大分子的应用中较为成功,但其中 QM 所能计算的区域还较小,大多数 QM/MM 方法是 MO/MM 或 VB/MM 模型。DFT/MM 计算方法还未见报道。从理论上来说,DFT/MM、ab initio 赝势法/MM 都是可行的,但现在计算方法还是一个空白。VB/MM(或 EVB)要优化参数和选择重要的共

振结构，全 VB 计算目前还不可能。许多人都在做 VB 与 MO 统一的工作，如能成功，将是很有意义的。

（3） 溶剂效应

溶剂效应研究中现有的模型还不够精确，发展溶剂效应的模型是一个可研究的领域。另外，现有的 MM、MD、FEP 的力场参数都由气态分子发展起来，直接应用于溶液中分子这样的近似计算，都会带来较大的，甚至很大的误差，因此有必要发展溶液中的分子力场参数。

（4） 并行算法的应用

随着并行计算机在我国的引进，发展生物大分子理论计算的并行算法是一项有意义的工作，同时，将几台单 CPU 的工作站联在一起也可实现并行算法，这样就可以在较短的时间内实现大体系高精度的计算。

2. 应用研究

（1） 酶抑制剂和受体调控剂的研究

许多蛋白质是催化体内某一生化反应的酶，或是小分子配体作用的受体。这些酶的抑制剂与受体调控剂（阻断剂或激动剂）往往可能成为很有价值的药物。从酶和受体的三维结构出发，根据配体小分子与受体（酶）活性部位在空间与化学上相互匹配及互补的原则，采用三维结构数据库搜寻和药物从头设计方法，可以发展新的酶抑制剂或受体调控剂。

（2） 离子通道抑制剂研究

人体内存在多种由蛋白质分子构成的离子通道，它们类似于活化酶，参与调节多种生理功能。钾、钙、钠等离子的选择性通透与心、脑血管与神经等系统的功能有密切关系。因此，离子通道的调控剂是一类重要药物，例如钠通道阻滞剂利多卡因是局麻和抗心律失常药物；钙通道阻滞剂硝苯啶、维拉帕米及地尔硫卓是抗高血压、心律失常和心绞痛的有效药物；钾通道启开剂 pinacidil 和 lemakalim 已试作降压药；钾通道阻滞剂 dotalol 和 UK68798 等则具有 III 类抗心律失常作用。通过研究构成离子通道的蛋白质分子的结构与功能，有可能设计出新的离子通道调节剂药物。国外已有一些实验室根据离子通道的三维结构，研究离子选择性通透的机理，但国内这方面研究尚未见开展，应予加强。

（3） 与基因有关的药物设计研究

人类许多疾病与基因异常密切相关。现已发现的癌基因有 100 多个，

已分离和肯定的抑癌基因也有10个左右。这方面的研究已引起国内外的广泛关注。蛋白质结构基础上的药物分子设计可以在这一研究领域中发挥重要作用。

抑癌基因P53是近几年来癌基因和抑癌基因研究的热点,它的突变在人类肿瘤中最常见。据报道,半数以上的结肠癌、乳腺癌、肺癌等均有P53基因的突变和缺失。突变后的P53不仅丧失抑癌活性,反而具有促进肿瘤向恶性转化的活性,也就是P53由抑癌基因转变为癌基因。医学界已经提出:如能针对突变后的P53基因设计相应的阻断剂,即可遏制其作为癌基因的作用。*Science*已经报道了P53的三维结构测定结果,为设计这类阻断剂提供了非常好的条件。因此这是一个非常值得注意的研究课题。

(4) 受体生物大分子结构测定及基于结构的药物设计

进行新型、高效、长效和无副作用的抗生育药物的计算机辅助设计,具有重要的意义。通过克隆、蛋白质表达等生物技术,分离纯化甾体激素受体,然后用结构测定技术(X衍射和MD-NMR等方法)结合计算机辅助结构解析方法,可以测定甾体激素受体亚型的三维结构。在此基础上用CADD方法设计新型的抗生育药物。

参 考 文 献

[1] 刘次全.量子生物学及其应用.北京:高等教育出版社,1990

[2] Goodfellow J M, Moss D S. Computer Modelling of Biomolecular Process. New York: Ellis Horwood, 1992

[3] Warshel A. Computer Modeling of Chemical Reactions in Enzymes and Solutions. New York: John Wiley & Sons, 1991

[4] Lybrand T P. *Curr Opin Struct Biol*, 1995, **5**: 224

[5] Greer J, Erickson J W, Baldwin J J, Varney M D. *J Med Chem*, 1994, **37**(8): 1035

[6] Gaussian 92/DFT. [Computer Program]. Gaussian, Inc, Pittsburgh, USA, 1993

[7] Malhotra A, Gabb H A, Harvey S C. *Curr Opin Struct Biol*, 1993, **3**: 24

[8] Elber R. *Curr Opin Struct Biol*, 1993, **3**: 260

[9] Kollman A. *Chem Rev*, 1993, **93**: 2395

[10] Field M J, Bash P A, *et al.* *J Compt Chem*, 1990, **11**: 700

[11] Allinger N L. *J Am Chem Soc*, 1977, **99**: 8127

[12] Allinger N L, Yuh Y H, Li J H. *J Am Chem Soc*, 1989, **111**: 8551

[13] Weiner S J, Kollaman P A, Case D A, Singh U C, Ghio C, Alagonag, *et al.* *J Am Chem

Soc, 1984, **106**: 765

[14] Weiner S J, Kollaman P A, Nguyen D J. *J Comput Chem*, 1986, **7**: 230

[15] Nilsson L, Karplus M. *J Comput Chem*, 1986, **7**(5): 591

[16] Brooks B R, Brucoleri R E, Olafson B D, States D J, Swaminathan S, Karplus M. *J Comput Chem*, 1983, **4**: 187

[17] Clark M, Cramer Ⅲ R D, Opdenbosh N. *J Comput Chem*, 1989, **10**: 982

[18] Vinter J G, Davis A, Saunders M R. *J Comp-Aided Mol Des*, 1987, **1**: 31

[19] Vedani A. *J Comput Chem*, 1988, **9**(3): 269

[20] Vedani A, Dobler M, Dunitz J D. *J Comput Chem*, 1986, **7**: 701

[21] Vedani A, Huhata D W. *J Am Chem Soc*, 1990, **112**: 4759

[22] Rappe A K, Casewit C J, Colwell K S, Goddard W A, Skiff W M. *J Am Chem Soc*, 1992, **114**: 10024

[23] Casewit C J, Colwell K S, Rappe A K. *J Am Chem Soc*, 1992, **114**: 10046

[24] Casewit C J, Colwell K S, Rappe A K. *J Am Chem Soc*, 1992, **114**: 10035

[25] Rappe A K, Colwell K S, Casewit C J. *J Inorg Chem*, 1993, **32**: 3438

[26] Mayo S L, Olafson B D, Goddard W A. *J Phys Chem*, 1990, **94**: 8897

[27] Van Gunsteren W F, Karplus M. *J Comput Chem*, 1980, **1**: 260

[28] Van Gunsteren W F, Benedsen H J C. *In Molecular Dynamics and Protein Structure*. Eds, J Hermans, Polycrystal Book Service. Westen Springs, IL, 1985

[29] Warshel A, Russell S T. *Quant Rev Biophys*, 1984, **17**: 283

[30] Books Ⅲ CL. *Curr Opin Struct Biol*, 1995, **5**: 211

[31] Hehre W J, Radom L, Schleyer P V R, Pople J A. *ab initio* Molecular Orbital Theory. New York: John Wiley & Sons, 1985

[32] SPARTAN. [Computer Program]. Wavefunction. Inc, Califirnia, USA, 1994

[33] Turbermol. [Computer Program]. Biosym Technologies, San Diego, USA, 1993

[34] Stewart J J P. *J Compt Aided Mol Des*, 1990, **4**(1): 1

[35] MOPAC. [Computer Program]. QCPE, Indiana University

[36] ZINDO. [Computer Program]. QCPE. Indiana University

[37] Parr G, Yang W. Density-Functional Theory of Atoms and Molecules. New York: Oxford University Press, 1989

[38] Thomas H. *Proc Camb Phil Soc*, 1927, **23**: 545

[39] Fermi R. *Accad Lincei*, 1927, **6**: 602

[40] Hohenberg P, Kohn W. *Phys Rev*, 1964, **136**: B864

[41] Kohn W, Sham L J. *Phys Rev*, 1965, **140**: A1133

[42] Payne C, Teter M P, *et al*. *Rev Mod Phys*, 1992, **64**: 1045

[43] Car R, Parrinello M. *Phys Rev Lett*, 1985, **55**: 2471

[44] Teter P, Payne M C, Allan D C. *Phys Rev*, 1989, **B40**: 12255

[45] Gillan J. Condens Matter *J Phys*, 1989, **1**: 689

[46] Aqvist J, Warshel A. *Chem Rev*, 1993, **93**: 2523

[47] Gresh N, Claverle P, Pullman A. *Theoret Chim Acta*, 1984, **20**: 1

[48] Richards G, King P M, Reynolds C A. *Protein Eng*, 1989, **2**: 319

[49] Barclay M, Butler J A V. *Trans Farad Soc*, 1938, **34**: 1445

[50] Nozaki V, Tanford C. *J Biol Chem*, 1971, **246**: 221
[51] Frommel C. *Theor Biol*, 1984, **111**: 247
[52] 蒋华良,陈凯先,嵇汝运. 化学通报,1995,(6):1
[53] Sinanoglu O, Abdulnar S. *Fed Proc*, *Suppl No 15*, 1965,**24**(2),Pt3: 12
[54] Klopman G, Andreozzi P. *Theoret Chim Acta*, 1980, **55**: 77
[55] Still C, Tempczak A, Hawley R C, *et al.* *J Am Chem Soc*, 1990, **112**: 6127
[56] Cramer J, Truhlar D G. *J Am Chem Soc*, 1991, **113**: 8305
[57] Cramer J, Truhlar D G. *J Comp Aided Mol Des*, 1992, **6**: 69
[58] Cramer J, Truhlar D G. *Science*, 1992, **256**: 213
[59] Pullman A, Berthod H. *Theoret Chim Acta*, 1975, **36**: 317
[60] Pullman A, *et al.* Quantum Theory of Chemical Reaction, Vol 2, Reidel, Dordrecht: 1
[61] Sullivan M. *Today's Chemist at Work*, 1995, **4**: 50
[62] Marshall R, Barry C D, *et al.* In: Computer-Asisted Drug Design, ACS Symposium Series 112, Olson E C, Christoffersen R E, Eds., Washington D. C., *American Chemical Society*, 1979, **205**
[63] Struthers R S, Rivier J, Hagler A T. *Ann N Y Acad Sci*, 1985, **439**: 81
[64] Andrea T A, Dietrich S W, *et al.* *J Med Chem*, 1979, **22**: 221
[65] Pettersson, Liljefors T. *J Comp Aided Mol Des*, 1987, **1**: 143
[66] Blaney J M, Dixon J S. *Persp Drug Disco Des*, 1993, **1**: 301
[67] Blaney J M, Weiner P K, *et al.* *J Am Chem Soc*, 1982, **104**: 6424
[68] Darden T, Mc Kinney J, *et al.* *Ann N Y Acad Sci*, 1986, **482**: 249
[69] Kuyper L, Davis S, LeBlanc H. In: Advances in Gene Technology: Protein Engineering and Production, ICSU Short Reports. Edt. Brew K. Oxford: IRL Press, 1988, **8**: 190
[70] Caldwell J, Kollman P. *Biopolymers*, 1986, **25**: 249
[71] Patel D J. *Proc Natl Acad Sci USA*, 1982, **79**: 6424
[72] Dicapua F A, Swaminathan S. *J Am Chem Soc*, 1991, **113**: 6145
[73] Sundaralingam M, Sekharudu Y C. *Science*, 1989, **244**: 1333
[74] Tirado-Rives J, Jorgensen W I. *Biochemistry*, 1991, **30**: 3864
[75] Wong C F, McCammon J A. *J Am Chem Soc*, 1986, **108**: 3830
[76] Bash P A, Singh U C, Langridge R, Kollman P A. *Science*, 1987, **236**: 564
[77] Merz K M, Kollman P A. *J Am Chem Soc*, 1989, **111**: 5649
[78] Morgan B P, Scholtz J M, Ballinger M D, *et al.* *J Am Chem Soc*, 1991, **113**: 297
[79] Merz K M, Murcko M A, Kollman P A. *J Am Chem Soc*, 1991, **113**: 4484
[80] Menziani M C, Reynolds C A, Richards W G. *J Chem Soc*, *Chem Commun*, 1989:853
[81] Hoops S C, Anderson K W, Merz K M. *J Am Chem Soc*, 1991, **113**: 8262
[82] Rao B G, Singh U C. *J Am Chem Soc*, **113**: 6735
[83] Reddy M R, Viswanadhan V N, Weinstein J N. *Proc Natl Acad Sci USA*, 1991, **88**: 10287
[84] Reddy M R, Kalish V, Palmer C, *et al.* Paper presented at Spring. 1993: Aes meeting in Denver
[85] Ferguson D M, Radmer R J, Kollman P A. *J Med Chem*, 1991, **34**: 2654

[86] Tropsha A, Hermans J. *J Protein Eng*, 1992, **5**: 29
[87] Cieplak P, Kollman P A. *J Comp Aided Mol Des*, 1993, **7**: 291
[88] Cummins P L, Ramnarayan K, Singh U C, *et al.* *J Am Chem Soc*, 1991, **113**: 8247
[89] Fleischman S H, Brooks C L. *Proteins*, 1990, **7**: 52
[90] Brooks C L, Fleischman S H. *J Am Chem Soc*, 1990, **112**: 3307
[91] McDonald J J, Brooks C L. *J Am Chem Soc*, 1991, **113**: 2295
[92] McDonald J J, Brooks C L. *J Am Chem Soc*, 1992, **114**: 2062
[93] Caldwell J W, Agard D A, Kollman P A. *Proteins*, 1991, **10**: 140
[94] Hirono S, Kollman P A. *J Mol Biol*, 1990, **212**: 197
[95] Rao S N, Singh U C, Bash P A, Kollman P A. *Nature*, 1987, **328**: 551
[96] Hwang J K, Warshel A. *Biochemistry*, 1987, **26**: 2669
[97] Mizushima N, Spellmeyer D, *et al.* *J Biol Chem*, 1991, **266**: 11801
[98] Pearlman D A, Kollman P A. *Biopolymers*, 1990, **29**: 1193
[99] Dang L X, Merz K M, Kollman P A. *J Am Chem Soc*, 1989, **111**: 8505
[100] Eriksson A E, Jones T A, Liljas A. *Proteins*, 1988, **4**: 274
[101] Maren T H. *Drug Dev Res*, 1987, **10**: 255
[102] Merz Jr K M. *Curr Opin Struct Biol*, 1993, **3**: 234
[103] Liang J Y, Lipscomb W N. *Biochemistry*, 1987, **26**: 5293
[104] Liang J Y, Lipscomb W N. *Biochemistry*, 1988, **27**: 8676
[105] Liang J Y, Lipscomb W N. *Int J Quantum Chem*, 1989, **36**: 299
[106] Merz K M, Hoffmann R, Dewar M J S. *J Am Chem Soc*, 1989, **111**: 5636
[107] Zheng Y J, Merz K M. *J Am Chem Soc*, 1992, **114**: 10498
[108] Jacob O, Cardenas R, Tapia O. *J Am Chem Soc*, 1989, **111**: 8692
[109] Sola M, Lledos A, Duran M, Berttran J. *J Am Chem Soc*, 1992, **114**: 869
[110] Krauss M, Garmer D R. *J Am Chem Soc*, 1991, **113**: 6425
[111] Head-Gordon T, Head-Gordon M, *et al.* *J Am Chem Soc*, 1991, **113**: 5989
[112] Gould I R, Kollman P A. *J Phys Chem*, 1992, **96**: 9255
[113] Böhm H J, Brode S. *J Am Chem Soc*, 1991, **113**: 5989
[114] Frey R F, Newton J, *et al.* *J Am Chem Soc*, 1992, **114**: 5369
[115] Weiner S J, Singh V C, *et al.* *J Am Chem Soc*, 1984, **106**: 6243
[116] Roterman L K, Lambert M H, *et al.* *J Biomol Struct Dyn*, 1989, **7**: 421
[117] Waszkowycz B, Hillier I H, Gensmanted N, *et al.* *J Chem Soc, Perkin Trans*, 1991, **2**: 225
[118] Waszkowycz B, Hillier I H, Gensmanted N, *et al.* *J Chem Soc, Perkin Trans*, 1991, **2**: 2025
[119] Brooks B R, Burecoleri R E, *et al.* *J Comp Chem*, 1983, **4**: 187
[120] Knowles J R, Albery W J. *Acc Chem Rev*, 1977, **10**: 105
[121] Aqvist J, Warshel A. *J Mol Biol*, 1992, **224**: 7
[122] Aqvist J, Fothergill M, Warshel A. *J Am Chem Soc*, 1993, **115**: 631
[123] Warshel A, Weiss R M. *J Am Chem Soc*, 1982, **102**: 6218
[124] Beddell C R. *The Design of Drugs to Macromolecular Targets*, New York: Wiley & Sons, 1992

[125] Gresh N, Rene B. *J Biomol Struct Dynamics*, 1994, **12**: 91
[126] Gresh N, Walter J, *et al*. *J Comp Chem*, 1995, **16**: 843
[127] Gresh N. *J Comp Chem*, 1995, **16**: 856
[128] Tapparelli C, Meternich R, *et al*. *Trends Pharmacol Sci*, 1993, **14**: 366
[129] Bode W, Mayr I, *et al*. *EMBO J*, 1989, **8**: 3467
[130] 蒋华良,陈凯先,唐赟,陈建忠,王沁泌,嵇汝运. 中国药理学报,1997, **18**(1):36
[131] Jiang H L, Chen K X, Tang Y. *The Forth China - Japan Joint Symposium of Drug Design and Development*. Xi'an, 1995:13
[132] Chen K X, Cresh N, Pullman B. *J Biomo Struct Dyn*, 1985, **445**
[133] Chen K X, Cresh N, Pullman B. *Nucleic Acids Res*, 1986, **14**: 1251
[134] 甘玲. 硕士论文,中国科学院上海药物研究所,1991
[135] 陈建忠,陈凯先,蒋华良,林茂伟,嵇汝运. 自然科学进展,1997,**7**:194
[136] 石根斌. 博士论文,中国科学院上海药物研究所,1995
[137] Sibley D R, *et al*. *Trends Pharmacol Sci*, 1992, **13**: 61
[138] Jin G Z, *et al*. *Sci Sin*(B), 1986, **29**: 527
[139] Tang Y, Chen K X, Jiang H L, *et al*. *Acta Pharmacol Sin*, 1996, **17**(1): 8
[140] Tang Y, Chen K X, Jiang H L, *et al*. *Chin Chem Lett*, 1996, **7**(3): 245
[141] Henderson R, *et al*. *J Mol Biol*, 1990, **213**: 899
[142] Mansour A, *et al*. *Eur J Pharmacol*, 1992, **227**: 205
[143] Hunter C A. *Chem Soc Rev*, 1994, **23**: 101

(陈凯先 蒋华良 唐 赟)

§ 2.7 同源蛋白质三级结构的预测

本世纪自然科学最伟大的事件之一是物理学与生物学结合而产生的分子生物学。它的标志是:①X 射线晶体学导致 DNA 分子双螺旋结构的发现,从而奠定了生命遗传的物理基础;②蛋白质折叠与生命运动关系的发现,从而奠定了生命活动的物理基础。遗传信息存储在核酸中,表达为蛋白质的一级氨基酸序列。在合成后,蛋白质便自发折叠成一个精确的三级结构。只有在折叠过程完成后蛋白质才能执行各种各样的功能,如催化、调控、化学输运、流动和结构支持等功能。1959 年的诺贝尔奖获得者 Anfinsen 证实蛋白质一级结构(氨基酸序列)唯一地决定它的三级结构(三维构象):"所有确定折叠过程及结果的信息均包含在氨基酸序列中。"人们把"蛋白质氨基酸序列决定其自然结构"称为生命的第二密码("DNA 序列决定氨基酸序列"称为生命的第一密码)。若能破译此密码,则现今的生物工程将出现空前突破。人们将能按照需要设计崭新的生化

催化剂、生理传感器、多肽药物、光化学能转换器、光机械能转换器以及纳米尺度上的能量和信息存储。

按照蛋白质折叠的热力学假设，完全折叠的蛋白质应处于其自由能全局最小的状态。这一假设曾经激励过许多一流的理论家，但是至今构造的经验自由能势函数的精度还不足以用于预测溶液中的蛋白质结构，当然巨大计算量和多重极小问题也是严重的理论障碍。自然折叠的蛋白质处在亚稳态，其折叠态和伸展态间的自由能纯差是个小项(21～42 kJ/mol)，只有高度精确的自由能势函数才能辨别预测结构的真伪。例如，图2.7.1把113个残基的蚯蚓血红蛋白(由4个α螺旋组成)和免疫球蛋白的变域(由β折叠组成)互换侧链，结果发现由目前的经验势函数决定的能量与自然蛋白的能量相差并不明显，真假难于辨认。

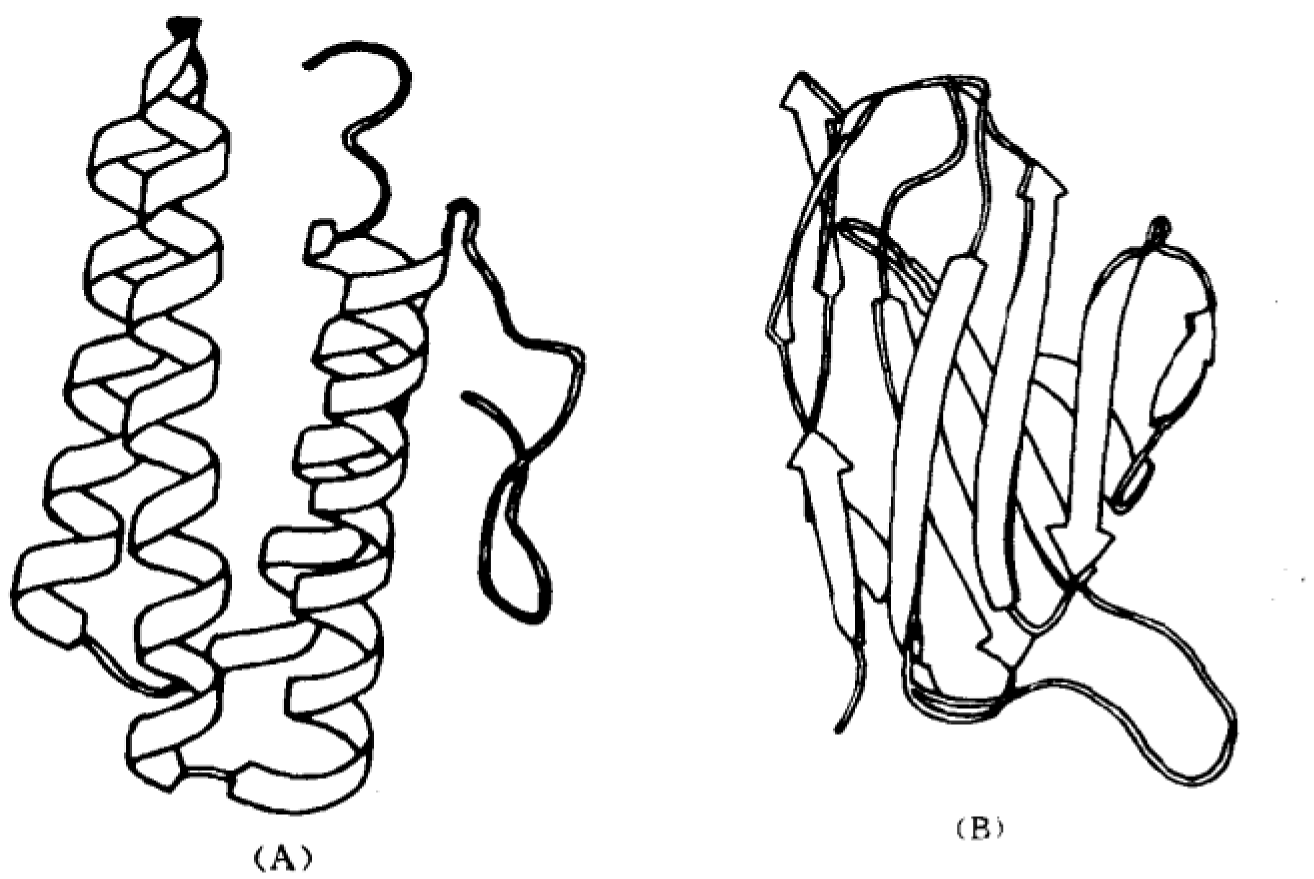


图　2.7.1

(A)由四螺旋束构成的蚯蚓血红蛋白；

(B)主要由β折叠组成的免疫球蛋白变区。

目前，蛋白质三级结构预测唯一可行的方法是蛋白质的同源模建，也就是从目标序列的同源结构来推断目标结构。它的直觉想法是同源蛋白质的结构常常是相似的。如果知道了一个或几个同源蛋白质的三级结构，把它们作为模板，可以推算出序列已知而结构未知的目标蛋白质的结构。

历史上第一个成功的例子是 Browne 等人(1969 年)根据同源溶菌酶(鸡卵清的)结构模建牛的 α 乳清蛋白的三级结构。20 年之后,Achasya 得到了高分辨(0.18 nm)的 α 乳清蛋白的晶体结构,经比较后发现,除了 C 端稍有偏差外,Browne 的模建结构是正确的。

2.7.1　同源模建的前途和问题

图 2.7.2 表示到 1994 年止测定蛋白质序列和结构的逐年增长的势头。事实上根据目前的报道,蛋白质序列的数目已超过 50000,而晶体结构已超过 3000。根据同源分析,基因组计划中待测序列的三分之一是与已测的序列同源相关的,而至少四分之一的已测蛋白质序列可在目前的 PDB(蛋白质结构库)中找到同源序列。由此,按目前已知的 120 个同源族计算,则自然界实际存在的同源蛋白族数目约是 1500。因为这里的分类延用了序列比较方法,其灵敏度限于序列同源性 30%以上,而事实上低序列同源性可能有高的结构同源性,例如硫-磷酸盐结合蛋白、铁传递蛋白和胆色素原脱氢酶之间没有明显的序列同源,但有相似的双叶阴离子结合结构。还有像 Rossmann 折叠、αβ 桶(TIM 桶,如磷酸丙糖异构酶)、β 薄卷饼,四螺旋束以及 β 免疫球蛋白,也常常出现在许多不同的蛋白质结构中。这样,经过折算,估计自然界中的同源蛋白族数约是 1 000。按照

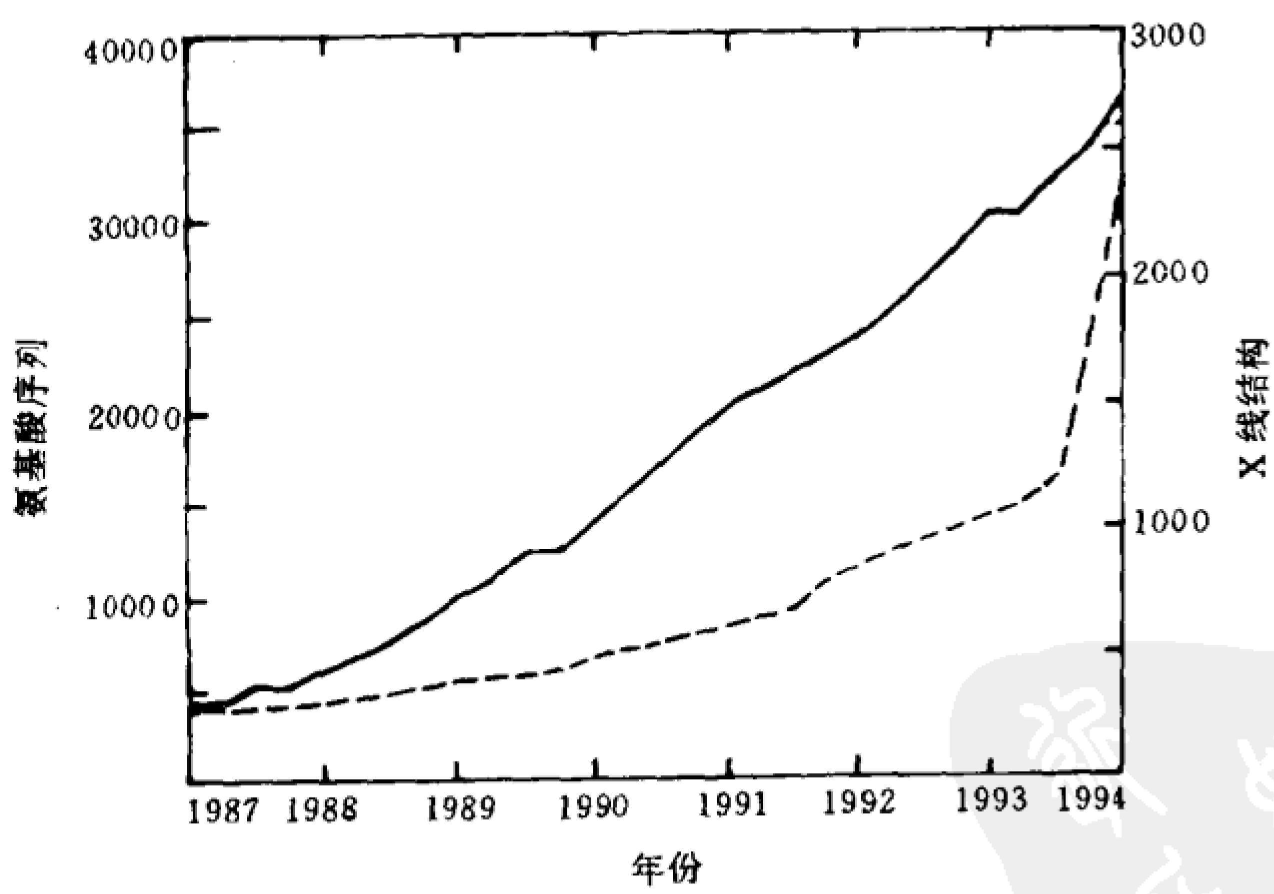


图 2.7.2　自 1987 年以来测定蛋白质序列和结构数目的增长情况

(引自 *Critical Reviews in Biochemistry and Molecular Biology*, 1995, **30**(1): 1)

目前X射线结晶图技术，核酸共振技术，特别是X射线同步光源技术的迅猛发展态势，完全有可能在下世纪初人类基因组DNA测序完成的同时，利用不断改进的同源结构预测方案把所有的自然蛋白质结构在原子分辨水平上描绘出来。人类将进入认识和驾驭生命世界的时代。

那么，同源蛋白的结构预测（或蛋白质的同源模建）已经发展到什么水平呢？这要从"何谓同源蛋白质"说起。"同源"是进化论中的名词，原指两个或两个以上由同一祖先蛋白质趋异进化而来的蛋白质。经典的进化学家检测同源性常从形态分类着手，而生物化学家常从种族和功能方面来考虑，因此常把蛋白质序列的相似性作为同源性的度量。现已明白这是两个既有联系又有区别的概念，因为会发现序列很不相似的两种蛋白仍属同一同源族，例如细菌丝氨酸蛋白酶与哺乳类动物丝氨酸蛋白酶。这表明，在进化历程中结构的保守性要比序列的保守性强。70年代结构生物学家Rossmann等人提出结构同源性概念，认为可用三级结构的相似性来鉴别同源性，这确实是一个好主意，但接着出现了许多复杂的现象。枯草杆菌蛋白酶与丝氨酸蛋白酶虽然整体折叠与序列各异，但均能水解肽链，活性部位相似，均有三联体H-S-D。肌动蛋白、热休克蛋白和己糖激酶在序列上约有10%的序列相似性，整体几何上偏差亦大，但具有相似的折叠拓扑，且具有相同的ATP结合位点。更有趣的是，扁桃酸消旋酶和粘液内酯化酶在芳香族酸降解代谢中是在不同反应步骤上催化，机制也不一样，但是结构很相似，序列很不相似（图2.7.3）。

这些现象说明了：①序列、结构和功能之间的定性和定量关系还不清楚。②序列相似性与结构相似性的关系是非线性的：序列很相似，结构会相似；序列不相似，结构可能会很相似。③结构很相似，序列可能很不相似。④结构相似要达到足够程度，才能有功能相似。⑤结构整体上不相似，但若有相似的功能Motif（序列上或结构的特征片段），还可能有相似的功能。上面讲到的枯草杆菌蛋白酶和丝氨酸蛋白酶的例子常称为趋同进化。它们并不来自同一祖先，在进化过程中两者的局部关联结构日趋接近，导致具有类似的功能。趋异进化和趋同进化在技术操作上是难于区别的。有人建议用"比较模建"代替"同源模建"。但由于历史缘由，目前大多仍沿用"同源模建"这个名称。

目前"同源模建"的方案基本上有两大类。一类是基于分段移植，即

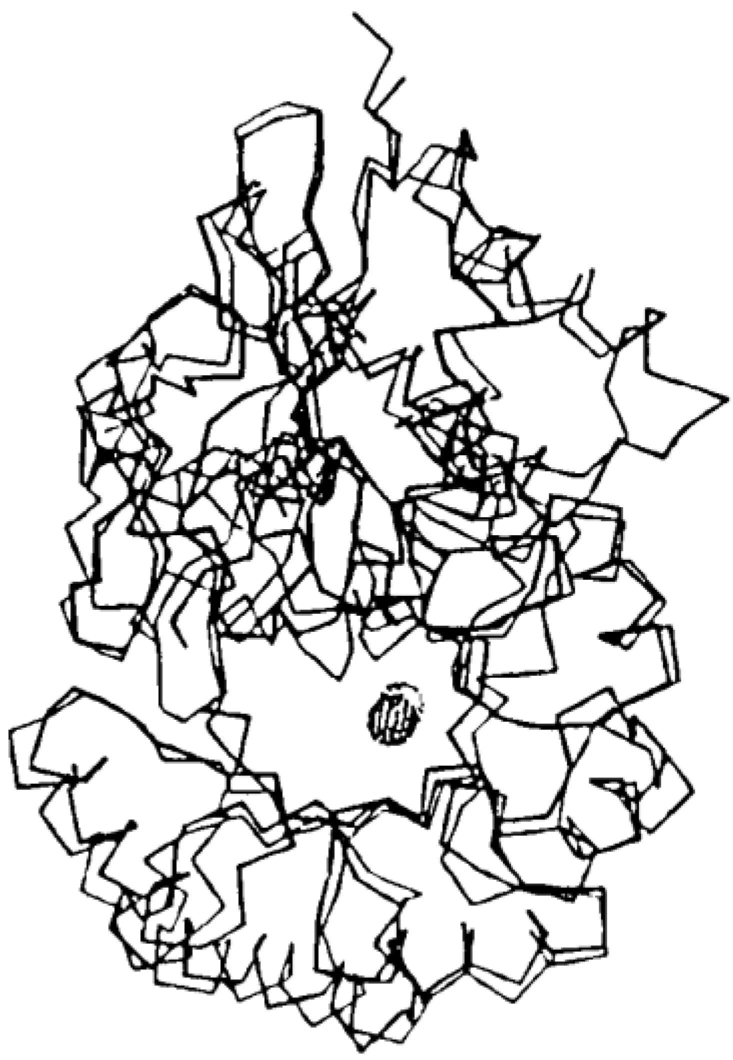


图 2.7.3　扁桃酸消旋酶与粘液内酯化酶的结构叠加

把已知同源结构的等价片段上的局部主链和侧链结构移植到目标蛋白的序列上。另一类是基于距离约束，即把同源结构中等价位置上的距离约束外推到目标蛋白的序列上，再按距离几何求定目标蛋白的原子坐标。当目标蛋白和同源蛋白的序列同源性大于 30%，这两种方案都能在原子分辨精度上模建或预测目标蛋白的三级结构。现在的问题是，如何在远缘同源(序列同源性低于 30%)时，仍达到原子分辨的预测精度。

2.7.2　我们的模建方案：PMODELING

通常的第一类模建方案可用图 2.7.4 表示。

设有一目标蛋白序列，同源模建它的三级结构，其步骤如下：

① 从数据库搜索同源参考蛋白；

② 从同源参考蛋白的结构信息确定它们的结构保守区(SCR)；

③ 从同源结构的 SCR 确定目标序列 SCR；

④ 在目标序列 SCR 上构建主链；

⑤ 用同源结构或构象搜索方法构建目标蛋白的结构可变区；

⑥ 由以上求得的目标蛋白的 C^{α} 原子坐标寻求目标主链的所有其他重、轻原子的坐标，并完成侧链安装；

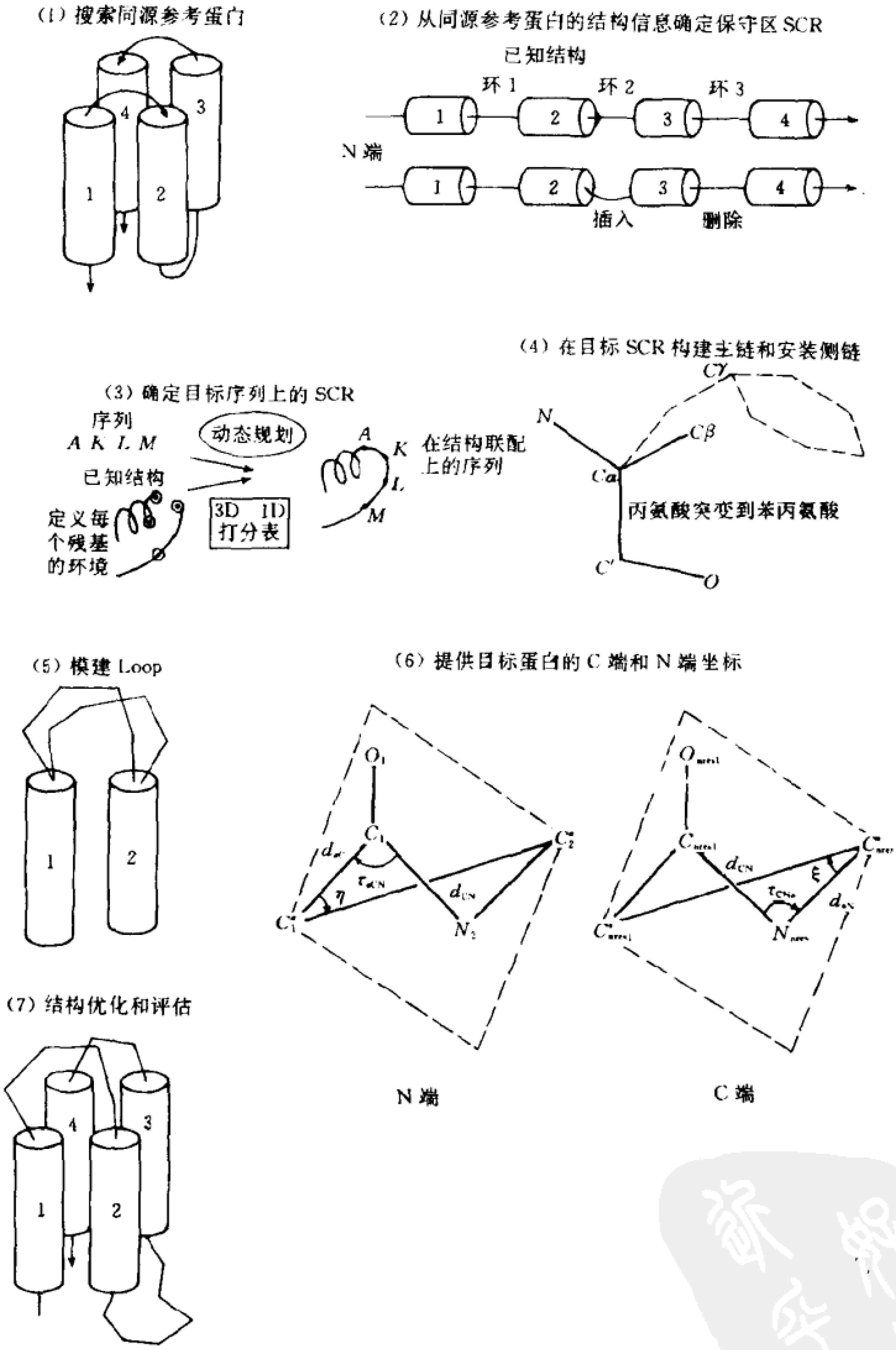


图 2.7.4 蛋白质的同源模建

⑦ 结构优化和评估。

目前已经商品化的模块软件有 HOMOLOGY 和 CONSENSUS (Biosym 公司),COMPOSER(Tripos 公司)。我们对“同源模建”的所有模块作了物理和数学的详细研究后[1~8],已经整合成蛋白质结构模建软件系统 PMODELING。图 2.7.5 展示 PMODELING 软件系统的方框流程图。

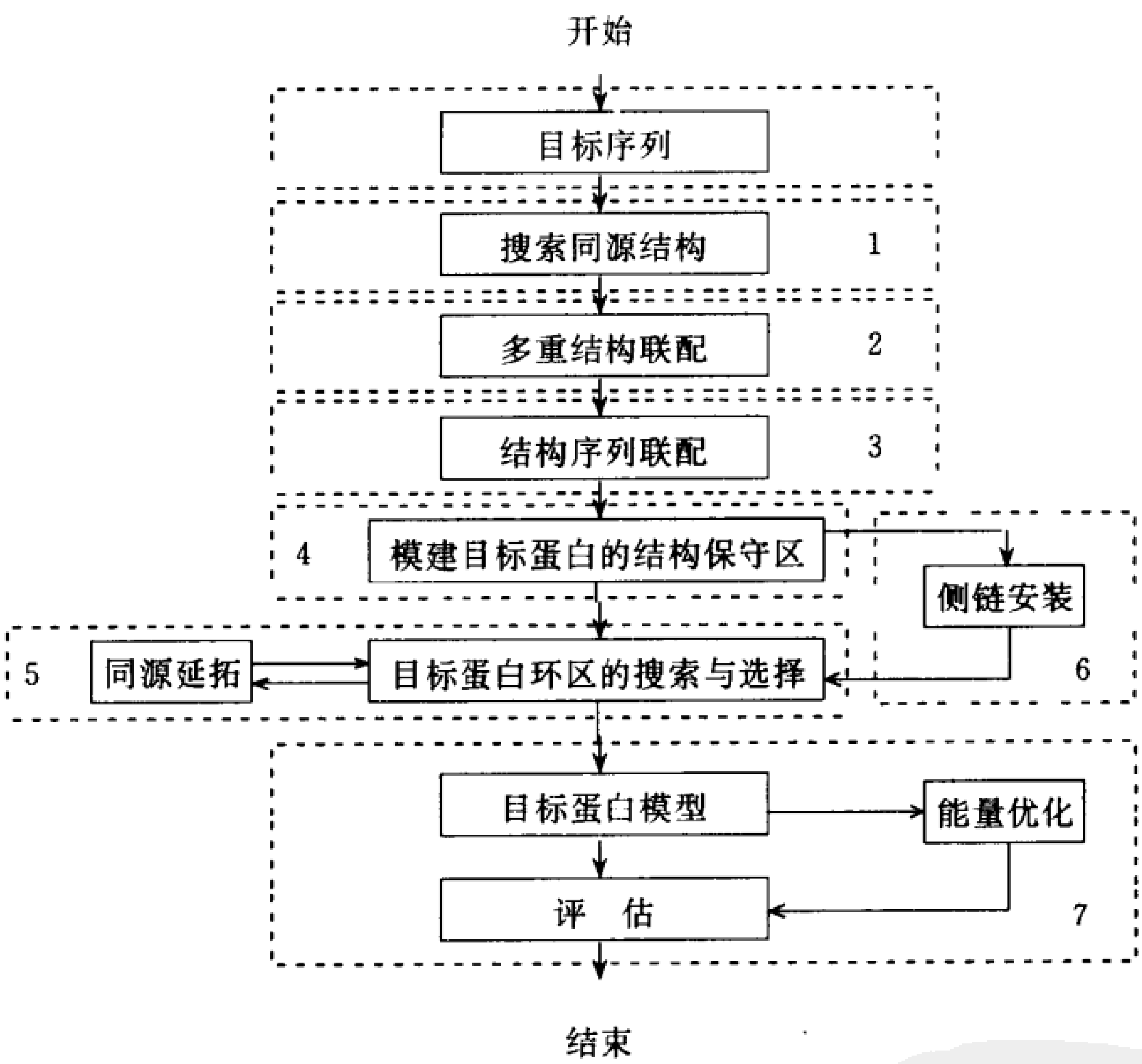


图 2.7.5 蛋白质三级结构模建系统的 PMODELING 流程图

2.7.3 蛋白质主链的模建

无论从图 2.7.4 或图 2.7.5 都可以感觉到,蛋白质结构的同源模建系统是规模不小的软件工程。在软件大国美国,也要 7 个人搞上 5 年才能推出商品软件。我们起初在 386 微机上设计,后来总算在“863”支持下有一台 SGI(R3000)图形工作站,先后有六七个人花了 4 年半时间,现已整

合成可供实验室使用的软件系统。它是我国第一个完全自行设计和研制的蛋白质结构同源模建的软件系统。这个系统有三大部分，共7个模块组成。这三大部分是：主链模建、侧链安装和结构评估。在主链模建部分要做的事情是：①根据手头的序列，要搜索到同源蛋白的结构，作为构建目标结构的模板。②把这些结构体比较分析（结构联配，structural alignment），以便确定这些模板蛋白组分氨基酸顺序间的拓扑等价关系，从而把同源结构域划分成结构保守区（SCR）和结构可变区（SVR）。③运用序列-结构比较把②中确定的拓扑等价关系投影到目标序列上。④通过模板结构的SCR和目标序列的SCR之间的对应关系设法用模板蛋白SCR C^α 原子坐标来确定目标序列SCR上的 C^α 原子坐标。⑤按同源移植和构象搜索方法构建目标蛋白的SVR C^α 原子坐标。⑥从这些已求定的 C^α 原子坐标采用自协几何约束条件，把目标主链的其他原子，如N、O、C和H的坐标也求导出来。这里不仅涉及到物理和数学上的深层问题，而且还涉及到若干计算复杂性问题[9]。

这里有两个理论方法问题：①蛋白质间的序列和/或结构分析。②目标蛋白质的SVR上模建问题。问题②是目前蛋白质工程中最有挑战性的问题。目前以郑强提出的标度松弛多拷贝解案为最佳[12]，但还有待系统地应用到模建问题中。我们也提出了一种解案，见文献[13]。又因为图2.7.5中的模块1、2、3都基于问题①的解决，所以在此仅谈①的理论问题。

既然同源模建基于结构相似性，那么，蛋白质间的序列和/或结构比较就成为自然的着手方法。这并不是一件容易的事情，例如图2.7.6，有两根折线，彼此的结点数不等。我们如何找到最多的等价（或对应）结点，而使在等价集上两根折线之间的根均方距离偏差达到极小。

1. 字符之间的比较

设有两个字符串

$$A = a_1 a_2 \cdots a_m,$$
$$B = b_1 b_2 \cdots b_n。$$

它们可以是DNA序列（由A、T、G、C四种核苷酸写成），也可以是氨基酸序列（由20种天然氨基酸写成），像

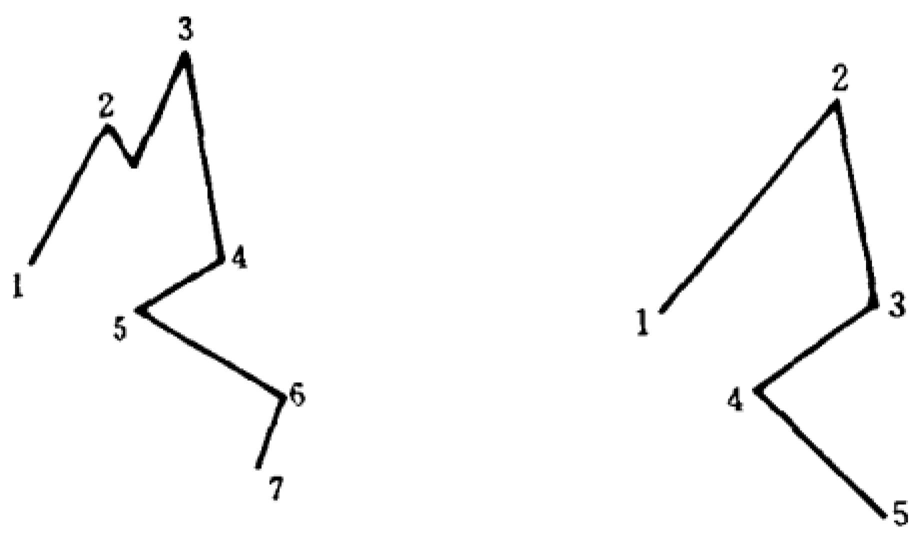


图 2.7.6 两条空间曲线的比较

$$\begin{array}{cccccc} a_1 & a_2 & a_3 & - & a_4 & a_5 \\ b_1 & b_2 & - & b_3 & b_4 & - \end{array}$$

称为字符串 $a_1a_2a_3a_4a_5$ 与字符串 $b_1b_2b_3b_4$ 间一种联配关系。像 a_1 对应 b_1 的关系称为置换，a_3 与空缺"－"的对应关系称为删去突变，而空缺"－"与 b_3 的对应关系称为插入突变。这三种突变是生物大分子进化中最主要的点突变。如果我们已经知道这三种序列之间突变的计量分数，就要问如何找到一种 A 序列突变到 B 序列的最佳方式(即找到 A 序列与 B 序列的联配)使得整个突变的计量分数达到最大。数学家证明，当 m、n 都超过 100 时，这种可能联配方式的总数是个天文数字，因此无法用穷举法找到最佳联配。然而数学家们已经找到一种方案，即动态规划的叠代格式：

$$M_{i,j}=\max\begin{cases} M_{i-1,j-1}+S_{i,j} \\ \max\limits_{1\leqslant k\leqslant i}(M_{i-k,j}-W(k,i,j)), \\ \max\limits_{1\leqslant l\leqslant i}(M_{i,j-l}-Z(l,i,j)) \end{cases} \tag{2.7.1}$$

其中 $M_{i,j}$ 表示子序列 $a_1a_2\cdots a_i$ 与子序列 $b_1b_2\cdots b_j$ 的最佳联配分数，$S_{i,j}$ 表示 a_i 置换成 b_j 的分数。(2.7.1)表示，若知道了(i,j)以前的子序列 $a_1a_2\cdots a_{i-1}$ 和子序列 $b_1b_2\cdots b_{j-1}$ 的最佳联配，则在(i,j)时最佳联配是下列三种可能方式中最好的一种：

第一种可能(置换)	第二种可能(k 个删去)	第三种可能(l 个插入)
$\cdots a_{i-l}\ a_i$	$\cdots a_{i-k}a_{i-k+1}\cdots a_i$	$\cdots a_i-\cdots-$
$\cdots b_{j-l}\ b_j$	$\cdots b_j-\cdots-$	$\cdots b_{j-l}b_{j-l+1}\cdots b_j$

在第二种和第三种可能方式中，相应的空缺罚分(即减少相似分数)是

$W(k,i,j)$和 $Z(l,i,j)$。当 W、Z 是逐个空缺位置上的罚分的线性叠加时，

$$W(k,i,j)=\eta^j_{0,i-k+1}v+\sum_{x=1}^{k}\eta^j_{i-k+x}u,$$
$$Z(l,i,j)=\zeta^i_{0,j-l+1}v+\sum_{y=1}^{l}\zeta^i_{j-i+y}u。\tag{2.7.2}$$

我们证实(2.7.1)可化为类 Gotoh 算法，使计算的时空复杂性大为减小：

$$M_{i,j}=\max(M_{i-1,j-1}+S_{i,j},F_{i,j},G_{i,j}),$$
$$F_{i,j}=\max(M_{i-1,j}-\eta^j_{0,i}v,F_{i-1,j})-\eta^j_iu,\tag{2.7.3}$$
$$G_{i,j}=\max(M_{i,j-1}-\zeta^i_{0,j}v,G_{i,j-1})-\zeta^i_ju。$$

叠代的初始条件是：

$$M_{0,0}=0;\ M_{i,0}=W(i,j,0);\ M_{0,j}=Z(i,0,j);$$
$$F_{0,j}=M_{0,j}-\eta^j_{0,i};\ G_{i,0}=M_{l,0}-\zeta^i_{0,j}。$$

当仅仅已知字符串的序列信息，而要找寻最佳联配，通称序列联配，分子生物学家常称为序列同源性分析，它广泛地应用在 DNA 序列和氨基酸序列比较分析中。例如，通过比较，可从一个序列的基因分布及其功能推断另一同源序列上的基因分布及其功能。在人类基因组计划中，要涉及到大尺度同源分析，即上述的 m、n 很大，常常是几万甚至十多万，即使采用上述改进方案也太慢了，常常要算上好几个星期(在 SUN 工作站上)。我们运用稀疏矩阵方法，可在数小时内较好地完成大尺度分析[14]。在序列联配时 $S_{i,j}$常取 Dayhoff 矩阵，这是 Dayhoff 从众多的先验蛋白质序列统计得到的，而罚分函数退化为简单形式：

$$W(k,i,j)=v+ku;\ Z(l,i,j)=v+lu。$$

v 称为起步罚分，而 u 称为单价罚分。另外在算法设计和编程技巧上还有许多窍门，此处不予细述。

这里要注意一点，上述(2.7.1)中的计量分数指相似分数，最佳的含义指相似分数最大。与此等价的表达方式是采用距离，即最佳的含义指描写两序列差异的距离最小。当用距离作计量分数时，上述算法的(2.7.1)和(2.7.3)形式不变，仅仅用min 替代max，用正号替代W 和Z 前的负号。

2. 蛋白质结构联配

上面已经说过两个蛋白质的序列不相似，并不能说明它们在三级结

构上不相似。因此，1976 年 Rossmann 等人提出用刚体叠加方法来找寻两个蛋白质的结构相似关系。把蛋白质结构看成连接各 C^{α} 原子的空间曲线，找寻一种刚体运动(即旋转加平移)，把两曲线叠加，使得在极大的 C^{α} 原子等价集上两曲线的根均方距离偏差(RMSD)达到极小。这个极其明确而又简单的问题至今仍未找到严格的数学解，但是已有 20 多种近似或经验的方法。后来发现，单纯的根均方距离偏差不一定是理想的最佳化目标函数，因为虽然 RMSD 体现了结构的整体约束力，但结构的形成还需要局部的物理与化学的约束力。于是 Rossmann 又在 RSMD 上加上局部折线走向偏差的约束。但是这个叠代方法太慢，且收敛性不好。90 年代出现一些新的改进方法，如 Taylor 等人的距离矩阵双动态规划方法[15]，以及 Blundell 等人的多重物化因子方法[16]。这些改进方法引入了新的物理思路：柔性的结构比刚性的结构叠加好。例如，珠蛋白家族(其中一个重要成员是血红蛋白)，若按纯刚性叠加方法，那么找到的拓扑等价位点是很少的，但从生物学家的直觉观察，它们是同源的，由几乎同样数目的 α 螺旋组成。它们虽在进化过程中经历了二级结构的局部运动甚至于变形，然而二级结构间的对应关系还是清楚的。

我们的物理直觉是，相似分数矩阵 $S_{i,j}$或距离分数矩阵 $D_{i,j}$是由三部分组成的：

① 序列信息部分 $W^1_{i,j}$；

② 局部构象部分 $W^2_{i,j}$；

③ 整体折叠部分 $W^3_{i,j}$；即

$$S_{i,j}\text{或 } D_{i,j}=\zeta W^1_{i,j}+\eta W^2_{i,j}+\lambda W^3_{i,j},$$

其中 ζ,η,λ 是相对权重。我们已经提出了两种构造分数矩阵的方法：一种是用统计方法构造 $S_{i,j}$，一种是用微分几何方法构造 $D_{i,j}$。两种方法可以达到同样好的效果。这里只讲第二种方法，第一种方法在张保红的博士论文“基于结构知识的蛋白质同源模建系统”中已详细描述。

设有两蛋白质的氨基酸序列为：

$$A = a_1a_2\cdots a_m,$$

$$B = b_1b_2\cdots b_n,$$

它们都是由 20 种氨基酸写成的，因为知道两蛋白质的结构，所以每一个氨基酸残基在空间中是确定的，不妨由它的 C^{α} 原子的坐标规定。因此 A

和 B 字符串的联配相当于有两条空间折线，每个结点上标有特定的氨基酸名，要找寻一种序元之间的最佳对应关系，使得在扣除空缺罚分后的距离分数达到极小。$D_{i,j}$中的 $W^1_{i,j}$可用前面说的 Dayhoff 矩阵来规定，而 $W^3_{i,j}$可用刚体叠加方法规定。通常的刚体叠加方法必须要借助于人的生物知识和图形视觉，无法自动化。我们已经提出了计算自动化问题的解决方案[2,8]，并采用微分几何方法确定 $W^2_{i,j}$。

图 2.7.7 表示由蛋白质的 C^α 原子连接成空间折线，它很好地刻划了蛋白质主链的空间走向。我们要用一种方法来描述它的几何特征。对这种不光滑的空间折线，若用样条逼近方法把空间折线光滑化，然后用经典微分几何处理，则此方法会影响最终的联配结果。1980 年 Rackovsky 和 Scheraga 找到了一种描述折线的自然方式，但它们要求此空间折线逐段相等，实际的 X 射线数据当然不是这样。Eisenmenger 于 1993 年提出一种正则化程序，可使 X 射线数据提供的空间折线满足 Scheraga 条件，但也要引进不小的误差和耗费很多 CPU 时间。我们的做法是，放弃空间折线逐步相等的假定，重新建立空间折线的微分几何(图 2.7.8)，在折线的每一个结点处建立一个局部坐标系(t_i, n_i, b_i)，即切向量、法向量和次法向量。它们的自然定义是：

$$t_i = \frac{v_i + v_{i+1}}{|v_i + v_{i+1}|},$$

$$n_i = p_i m_i = p_i \frac{v_i - v_{i+1}}{|v_i - v_{i+1}|},$$

$$p_i = \mathrm{sgn}(m_i \cdot n_{i-1}),$$

$$b_i = t_i \times n_i,$$

此处 $v_i = r_i - r_{i-1}$，而 r_i 是 C^α_i 原子的空间坐标$(i=1,2,\cdots,m)$。这种递归关系的初始条件是 $p_2=1$。所以，对于实际的 X 射线坐标

$$t_i \cdot n_i \neq 0,$$

类似经典微分几何，定义曲率 k_i 为：

$$k_i = q_i \frac{\cos^{-1}(t_i \cdot t_{i+1})}{|v_{i+1}|} = \frac{q_i \chi_i}{|v_{i+1}|},$$

$$q_i = p_i \cdot \mathrm{sgn}[(t_i \cdot t_{i+1}) \cdot b_{i+1}]。$$

而挠率 τ_i 为

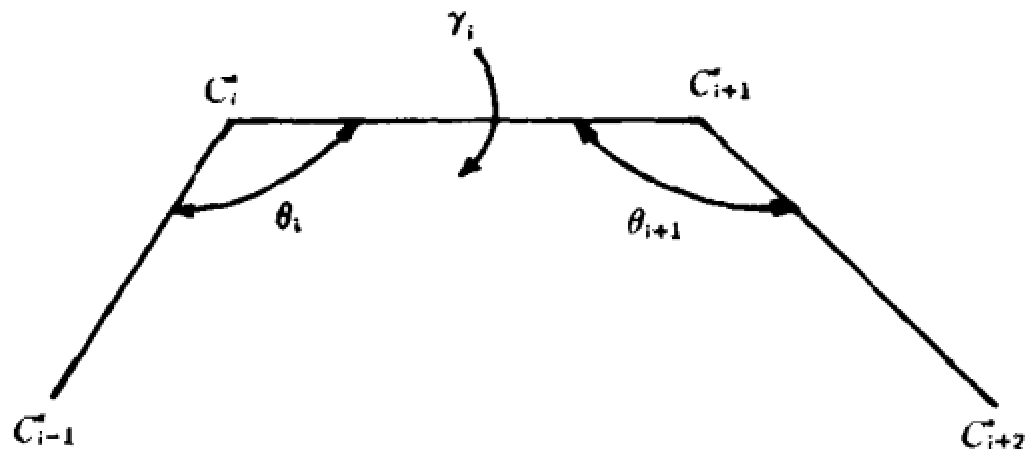


图 2.7.7 多肽链的虚键表示

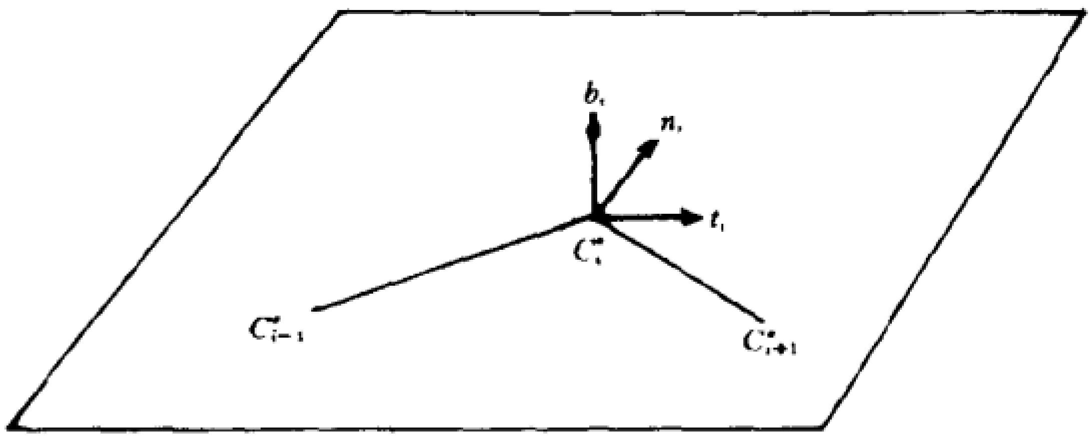


图 2.7.8 当 $p_i=+1$，在 C_i^a 处的局部标架

$$\tau_i = u_i \frac{\cos^{-1}(b_i \cdot b_{i+1})}{|v_{i+1}|} = \frac{u_i \cdot \gamma_i}{|v_{i+1}|},$$

$$u_i = \mathrm{sgn}(n_i \cdot b_{i+1})。$$

q_i 和 u_i 指示的符号约定是为了保证曲线的(κ,τ)表示和坐标表示之间的唯一关系。若给定$\{\kappa_i,\tau_i,i=2,3,\cdots,m-2\}$和第一个虚键角 θ_2(图2.7.7)，则 C_i^a 坐标$(i=1,2,\cdots,m)$唯一地被确定。我们的目的是找到描写两个蛋白质的局部构象差异的距离。设两蛋白质结构的曲/挠率表示为(κ_i,τ_i)和(κ_j',τ_j')。因为在(κ,τ)平面上有两种不连续现象，所以这个距离不可简单地用(κ,τ)空间中的欧氏距离来定义。第一种是，构象微小的变化会引起p_ip_{i+1}变号，因而(κ_i,τ_i)会经历不连续变化：

$$\begin{aligned} \kappa_i &\longrightarrow \bar{\kappa}_i = -\kappa_i, \\ \tau_i &\longrightarrow \bar{\tau}_i = -\left(\frac{\pi}{|v_{i+1}|} - |\tau_i|\right)\mathrm{sgn}(\tau_i)。 \end{aligned} \tag{2.7.4}$$

第二种不连续性要复杂一些。一般对给定的 κ_i 和 x_i 有两种可能的 C_{i+2}^a坐标，由 q_i 的正负号来区别。当$|\tau_i|$增加时这两个坐标彼此靠近。在极大的

$|\tau_i|$处，对给定的 x_i，这两个坐标点重合。这表示出现了不连续性，因为对应于 $q_i=+1$ 和 $q_i=-1$ 的这两个点在实际的构象中是重合的，但在(κ,τ)平面上被一个禁区分开了。我们已经证明这个禁区的边界由下式规定：

$$\kappa_i = q_i K(|\tau_i|, \theta_i, |v_i|, |v_{i+1}|), \tag{2.7.5}$$

此处

$$K(|\tau_i|, \theta_i, |v_i|, |v_{i+1}|) = \frac{\sin^{-1}[\sin\lambda_i \cdot \sin(|\tau_i|, |v_{i+1}|)]}{|v_{i+1}|}。\tag{2.7.6}$$

这里的 λ_i 是 t_i 和 v_i 的夹角且由下式确定

$$\sin\lambda_i = \frac{|v_i|\sin\theta_i}{\sqrt{|v_i|^2 + |v_{i+1}|^2 - 2|v_i| \cdot |v_{i+1}| \cdot \cos\theta_i}}。\tag{2.7.7}$$

现在来构造距离。为了消除 $p_i p_{i+1}$的不连续性，我们作一预处理，凡需要者均作变换(2.7.4)，使得待比较的(κ_i,τ_i)和(κ_j',τ_j')有同样的 $p_i p_{i+1}$值。这样，构造距离时只须考虑第二类不连续性。

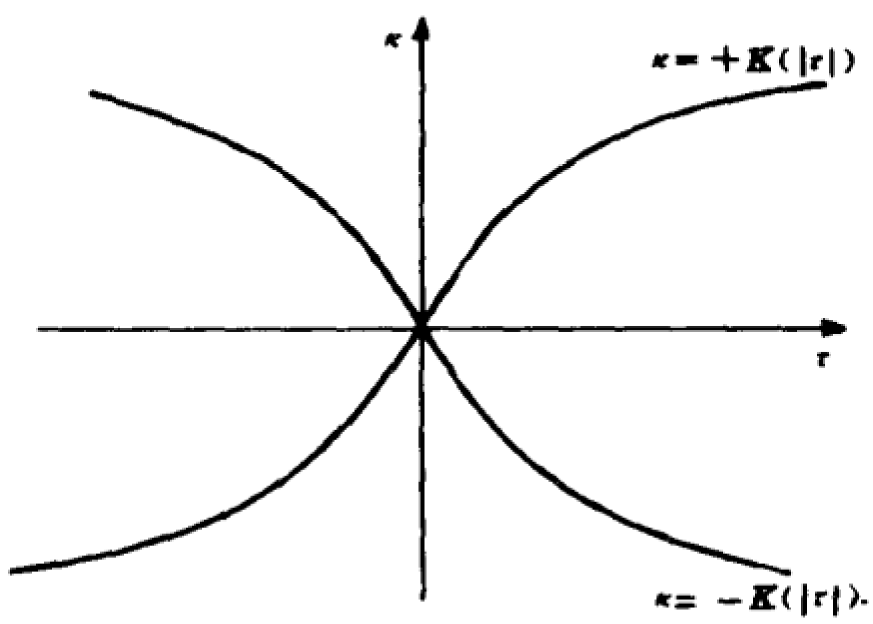


图 2.7.9 在(κ,τ)平面上的间断区

情况 1：$\text{sgn}(\kappa_i)=\text{sgn}(\kappa_j')$。

此时(κ_i,τ_i)和(κ_j',τ_j')处在 τ 轴的同一侧。

再考虑到 $\gamma_j'=\pm\pi$ 时，有一致的构象，即$\left(\kappa_j', \frac{u_j'\pi}{|v_{j+1}'|}\right)$和$\left(\kappa_j', \frac{-u_j'\pi}{|v_{j+1}'|}\right)$在构象上是等价的。所以距离函数 $\rho_{i,j}$应取

$$\rho_{i,j} = \min(\rho_{i,j}^1, \rho_{i,j}^2), \tag{2.7.8}$$

此处

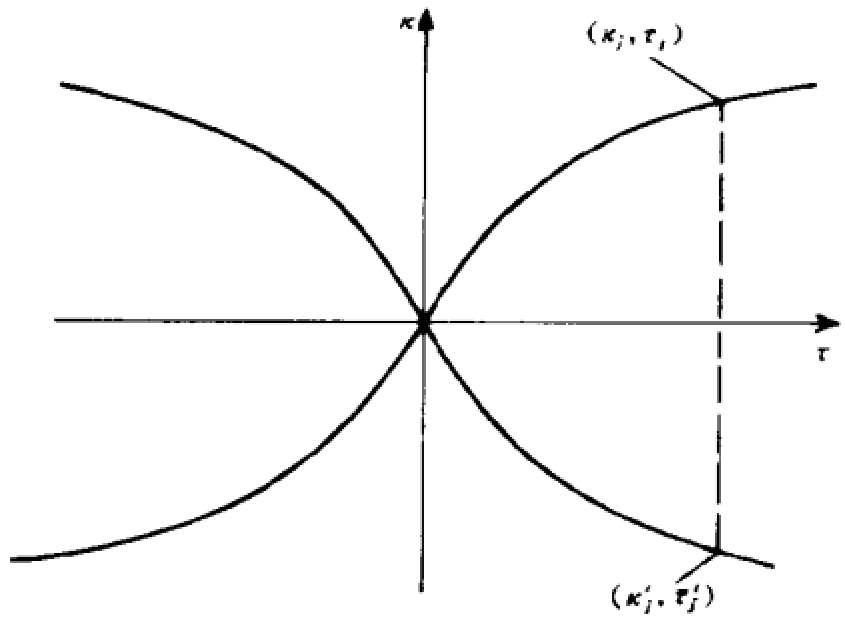


图 2.7.10 在(κ,τ)平面上，两点(κ_i,τ_i)和(κ_j',τ_j')之间的路径，其中 κ 值具有相反符号

$$\rho_{i,j}^1 = \sqrt{(\kappa_i - \kappa_j')^2 + (\tau_i - \tau_j')^2},$$

$$\rho_{i,j}^2 = \sqrt{(\kappa_i - \kappa_j')^2 + \left(\tau_i - \tau_j' + \frac{2u_j'\pi}{|v_{j+1}'|}\right)^2}。$$

情况 2：$\mathrm{sgn}(\kappa_i) = -\mathrm{sgn}(\kappa_j)$。

此时(κ_i,τ_i)和(κ_j',τ_j')的分布情况如图 2.7.10 所示。考虑到它们被禁区隔开，距离函数的自然定义为：

$$\rho_{i,j} = \min[\rho_{i,j}(\theta_i, |v_i|, |v_{i+1}|), \rho_{i,j}(\theta_j', |v_j'|, |v_{j+1}'|)],$$

此处

$$\begin{aligned}\rho_{i,j}(\theta,v,v_1) = \min_{\tau}\{&[(\kappa_i - K(|\tau|;\theta,v,v_1))^2 \\ &+ (\tau_i - \tau)^2]^{1/2} + [(\kappa_j' + K(|\tau|;\theta,v,v_1))^2 \\ &+ (\tau_j' - \tau)^2]^{1/2}\}。\end{aligned} \tag{2.7.9}$$

从(2.7.8)和(2.7.9)，我们确定了距离分数 $D_{i,j}$中第二部分。有了距离分数矩阵 $D_{i,j}$，我们容易制成结构联配软件CONTRAST。经上百族蛋白质结构的测试，此软件与国外同类型软件相比，其优点明显，具体的结果请看文献[1,6]。

3. 蛋白质序列-结构联配

序列-结构联配指联配成员中有一部分仅有序列信息。这种联配用处很多，例如数据库搜索、结构评估以及目标序列 SCR 的确定等。要确定目标序列 SCR，难度就在于目标蛋白仅知序列，如何从模板结构的 SCR 求

得。如果有了序列结构联配方法，就可解决问题。

蛋白质是由众多物理力（疏水相互作用、静电力、成链作用、氢键连接、范德瓦尔斯力等）稳定成形的，因此在一定结构中每个位置上某氨基酸所处的结构环境体现了它经受的物理力。将氨基酸残基所处的环境分成 64 类：4 类二级结构（$\alpha,\beta,+\varphi$，其他），2 类溶剂可及性（$>7\%$，$<7\%$），8 类氢键连接状态（侧链-侧链，侧链-主链 N 原子，侧链-主链 O 原子）。然后把上文(2)中的结构联配方法用于已知 110 族蛋白质结构（共 444 个蛋白质结构），得到 110 幅结构联配表，例如图 2.7.11 是三个磷酸酯酶家族成员的结构联配。从这些联配表，可以统计出处于某环境 s 的一种残基 i 突变到另一种残基 j 的频数 $N_{si,j}$，按信息值的概念[17]，定义相似分数矩阵：

```
    10        20        30        40        50        60        70
ALWQFNGMIKCKIPSSEPLLDFNNYGCYCGLGGSGTPVDDLDRCCQTHDNCYKQAKKLDSCKVLVDNPYT
ALWQFRSMIKCAIPGSHPLMDFNNYGCYCGLGGSGTPVDELDRCCETHDNCYRDAKNLDSCKFLVDNPYT
SLVQFETLIMKIA-GRSGLLWYSAYGCYCGWGGHGLPQDATDRCCFVHDCCYGKATD···C······NPKT
 hhhhhhhhhh                           hhhhhhhhhhhhh

    80        90        100       110       120       130
NNYSYSCSNNEITCSSENNACEAFICNCDRNAAICFSKV-P-YNKEHKNLD-KKNC······lbp2-
ESYSYSCSNTEITCNSKNNACEAFICNCDRNAAICFSKA-P-YNKEHKNLDTKKYC······lp2p-
VSYTYSEENGEIICG-GDDPCGTQICECDKAAAICFRDNIPSYDNKYWLFP-PKDCREEPEPC lpp2R
  bbb      bbb     hhhhhhhhhhhhhhhhhh              hhh
```

图 2.7.11 三种磷酸酯酶 A_2 蛋白质的结构联配

标识符 h 表示 α 螺旋，b 表示 β 折叠。

```
·········IIGGVESIPHSRPYMAHLDIVTEKCLRVICGGFLISRQFVLTAAHCKGR···EITVILGA
CGVPAIQPVLSIVNGEEAVPGSWPWQVSLQDKT···GFHFCGGSLINENWVVTAAHCGVTT-SDVVVAGE
·········VVGGTEAQRNSWPSQISLQYRSGSSWAHTCGGTLIRQNWVMTAAHCVDRELTFRVVVGE
·········IVGGYTCGANTVPYQVSLNS······GYHFCGGSLINSQWVVSAAHCYKS···GIQVRLGE
HDVRKRESTQQKIKVEKQIIHESYNSVP···NLHDIMLLKLEKKVELTPAVNVVPLPSPSDFIHPGAMCWA
FDQGSSSEKIQKLKIAKVFKNSKYHSLT···INNDITLLKLSTAASFSQTVSAVCLPSASDDFAAGTTCVT
HNLNQNNGTEQYVGVQKIVVHPYWNTDDVAAGYDIALLRLAQSVTLNSYVQLGVLPRAGTILANNSPCYI
DNINVVEGNEQFISASKSIVHPSYNSNT···LNNDIMLIKLKSAASLNSRVASISLPTS···CASAGTQCLI
AGWGKTGV-RDPTSYTL REVELRIMDEXOCV--DY-RYYEYKFQVCVGSPTTLRAAFMGDSGGPLLC---
TGWGLTRY-A-NTPDRLQQASLPLLSNTNCK--KYWGTKIKDAMTCAGASG--VSSCMGDSGGPLVCKKN
TGWGLTRT-NGQLAQTLQQAYLPTVDYAICSSSSYWGSTVKNSMVCAGGDGV-RSGCQGDSGGPL HCLVN
SGWGNTKSSGTSYPDVLKCIKAPISDSSCK-SAYPG-QITSNMFCAGYLEGGKDSCQGDSGGPVVC---
-AGVAHGIVSYGHPD--AKP--PAIFTRVSTYVPWINAVIN--     3rp2
GAWTLVGIVSWGSST--CSTSTPGVYARVTALVNWVQQTLAAN     4cha
GQYAVHGVTSFVSRLGCNVTRKPTVFTRVSAYISWINNVIASN     3est
-SGKLQGIVSWGSCC--AQKNKPGVYTKVCNYVSWIKQTIASN     2ptn
```

图 2.7.12 胰蛋白酶（2ptn）序列和三个丝氨酸蛋白酶（3rp2，4cha，3est）结构之间的序列-结构联配

$$S_{si,j}=\ln\frac{P_{si,j}}{P_{si}\cdot P_j}=\ln\frac{N_{si,j}/N}{(N_{si}/N)\cdot(N_j/N)},\qquad(2.7.10)$$

这里 P 表示概率，且

$$N_{si}=\sum_{j=1}^{21}N_{si,ji};\quad N_j=\sum_{s=1}^{64}\sum_{i=1}^{21}N_{si,j};\quad N=\sum_{j=1}^{21}N_j。$$

我们将胱氨酸和未形成二硫键的半胱氨酸当作两种氨基酸来处理，因此共有 21 种氨基酸。(2.7.10)是 64 张 21×21 的矩阵表。在运用动态规划方法(2.7.1)实行序列-结构联配时，空缺罚分函数中不计及目标蛋白的贡献[9]。下面给出一例，模板蛋白是巨细胞蛋白酶 II(3rp2)、α 糜蛋白酶(4cha)、胰弹性蛋白酶(3est)，同属丝氨酸蛋白酶家族，目标序列是胰蛋白酶。先把三个模板结构联配好，然后再与目标序列实行结构-序列联配(见图 2.7.12)。

上述字符串的联配是主链模建中的关键，有了它可以搜索同源结构，确定模板结构的 SCR 和 SVR，再投影到目标序列上。接下来主要是技术问题：①用无偏刚体叠加技术求定目标序列 SCR 上的 C^α 原子坐标；②用 SCR 延拓和构象搜索方法求定目标序列 SVR 上的 C^α 原子坐标；③用几何约束条件从 C^α 原子坐标求定主链上的其他原子坐标。这些技术细节可参看文献[9]。可告诉读者的是，经 20 多族同源蛋白的回顾性检测表明，PMODELING 比 HOMOLOGY 和 COMPOSER 好：操作自动化，模建精度高，运行时间少。而且，我们还对蝮蛇磷酯酶、细胞因子受体 CD40 的配体 gp39、小鼠巨细胞蛋白酶和人白介素 13 等至今未知结构的蛋白作了预测性模建，有待实测数据的检验。

2.7.4 侧链的安装

在蛋白质的主链确定后，侧链的堆积方式是由各个残基的侧链扭转角 χ 的组合所决定的。一般认为，其中能量最小的组合就是蛋白质自然折叠的侧链堆积方式。预测侧链构象通常遇到的困难是：①此种组合数太大。对一个 150 个氨基酸长度的蛋白质，侧链组合总数约为 10^{105}，因而要穷举是不大可能的。②目前用来计算能量的各种势场还不能达到预测所需的精度。

经蛋白质晶体结构的研究发现，侧链构象的扭转角 χ 常分布在一定范围内。Ponder 和 Rechasds[18]统计了 19 个高分辨晶体结构，得到 83 个最可几构象，称为转子。后来转子不断扩充和改进，最近 Dunbrack 等[19]经大量统计得到了主链骨架依赖的转子库。即使从如此有限的转子库中

找寻最合适的组合方式，也决非容易之事。最近 Desmet 等提出的死端排除定则[20,21]，希望在可接受的计算时间内严格寻找最低全局能量的侧链构象组合。但经我们大量测试，此方法在模板主链上安装侧链时要比侧链的同源移植引来更大的误差。于是我们提出一种同源移植与死端排除的组合方式，即尽可能采用模板侧链的信息，仅在必要时才用死端排除法对转子库进行搜索。结果发现效果较好[10]。目前，各种侧链安装方法都含有两个假设：①安装过程中，主链结构不受影响；②侧链转子取理想的最可几构象位置。最近的晶体结构数据表明，这两个假设必须重新审查。这些方法(包括我们的)的精度实际上已经达到了在这两个假设下可能达到的上限。面对这种挑战，我们正在筹划可行的方案。郑强与美国同事提出的标度松弛多拷贝思路和理论物理中的“轨道簇”或时间依赖 Hartree 技术都是值得关注的。

2.7.5 灵敏的蛋白质结构评估方法

早先，人们认为正确折叠与错误折叠的自由能差别应该是一种理想的判据。然而，Novotny[22]等人把一个全由 α 螺旋组成的蚯蚓血红蛋白和一个全由 β 折叠组成的免疫球蛋白变区互换侧链，在能量优化后发现，它们在真空状态下由经验势场决定的能量与自然蛋白质的能量相差并不明显，只是在错误折叠的蛋白质中，有较多的非极性基团暴露于环境以及较多的极性基团埋藏于内核。可见，目前的经验势场还不足以用来区别错误折叠和正确折叠。Baumann 等人[23]分析了 128 个晶体结构，然后与一些模建的蛋白质结构相比，发现有三个模建蛋白在内核和溶剂可及表面有异常的极性和非极性侧链的分布。Sander 研究组[24]利用从晶体结构统计获得的原子溶剂化倾向性参数评估模建结构，此方法对整体评估是有效的，但对于局部折叠的错误则难于辨认。Ooi 接触数(即进入某原子球内的残基数)作为残基接触的参数能够反映蛋白质的折叠特征。Topham[25]发现模建结构与晶体结构的 C^α 原子位置偏差与 C^α 原子的 Ooi 接触数的倒数有一致的趋势，Ooi 接触数大的位置偏差小，反之偏差则大，但它难于给出整体结构的评估。Bowie 等人[26]提出一种剖面分析的方法，用来识别折叠模式，他们用局部二级结构类型、溶剂可及性和被极性原子遮盖的侧链面积来刻划残基所处的环境，经组合有 18 类环境。由已知的蛋白质结构及其序列统计出残基类型对环境类型的倾向性分数。

Luthy[27]等人把此概念推广到蛋白质结构评估，即判别一序列与其结构的兼容性。这种方法已广泛应用，甚至 Biosym 公司由此推出软件产品 Profile-3D。但是我们发现，它对局部错误并不敏感，难于作出正确的判定。

其实，一个三维模建结构的评估可在几个结构层次上进行：①整体折叠的可靠性；②局部区域的可靠性；③立体化学参数的检查(键长、键角及氢键图等)。Morris 等人[28]研究了已知蛋白质结构，发现很多立体化学参数的分布具有一定的规律，例如，二面角 ϕ 和 ϕ 常满足 Ramachandran 图，Proline 处的 ϕ 常取负值，(倘若出现 $+\phi$，则在进化上异常保守)，C^{α}—N—C—C^{β} 二面角一般为 33.9°。PROCHK 软件[29]用这些参数检查结构的立体化学品质。

如何既能判断整体折叠的正确性，又能检测到局部折叠的错误呢？

在我们的蛋白质模建系统 PMODELING 中，评估模块 ASSESSOR 试图对此问题提出一种解决方案，经大量测验，ASSESSOR 与 Profile-3D 相比，评估品质更好，特别对局部错误折叠的识别更为灵敏。ASSESSOR 不仅适用于模建结构的评估，而且也可用于对晶体结构和 NMR 结构测定的评估。

1. ASSESSOR 的算法

(1) 残基-环境倾向性分数

前文我们已经求得了处于某环境 s 的一种残基 i 突变到另一种残基 j 的频数 $N_{si,j}$。那么，在环境 s 中出现 i 的频数为：

$$N_{si} = \sum_{j=1}^{21} N_{si,j}, \tag{2.7.11}$$

而环境 s 出现的频数和残基 i 出现的频数分别为：

$$N_s = \sum_{i=1}^{21}\sum_{j=1}^{21} N_{si,j}, \quad N_i = \sum_{s=1}^{64}\sum_{j=1}^{21} N_{si,j}, \tag{2.7.12}$$

于是残基 i 在环境 s 下的，环境 s 的以及残基 i 的出现概率分别是：

$$P(si) = \frac{N_{si}}{N}, \quad P(s) = \frac{N_s}{N}, \quad P(i) = \frac{N_i}{N}, \tag{2.7.13}$$

其中

$$N = \sum_{s=1}^{64}\sum_{i=1}^{21}\sum_{j=1}^{21} N_{si,j},$$

$$S_{si} = \ln \frac{P(si)}{P(s)P(i)}。 \tag{2.7.14}$$

图 2.7.13 列出了几种典型的例子，显示了不同局部环境下氨基酸的倾向性分数。

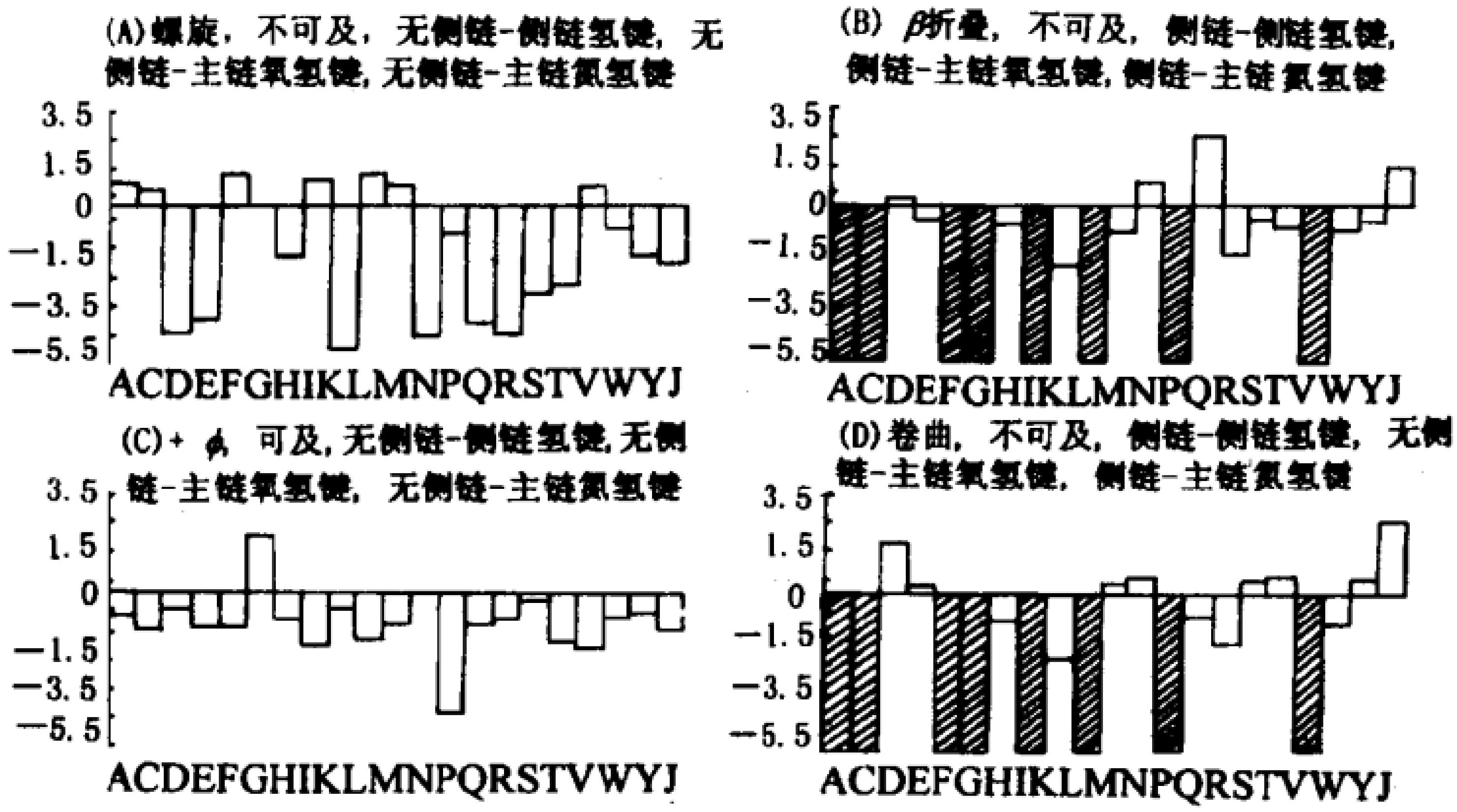


图 2.7.13 在不同局部环境下各种氨基酸的倾向性分数

黑条表示此氨基酸不可能在该环境下出现。

（2）错误结构的构建

为了确定结构判定的上下界，需要构建一些错误结构。

我们采用类似 Novotny[22]的方法，选取长度相同而二级结构差异较大的两种蛋白质，以互换侧链构建错误蛋白。例如长度为 113 个残基的蚯蚓血红蛋白(lhmd)A 链和免疫球蛋白的轻链(1mcp)，其中蚯蚓血红蛋白 A 链由 4 个 α 螺旋构成，免疫球蛋白轻链则由 β 折叠构成。互换两个结构的侧链，然后采用 Biosym 公司的 2.9 版 Discover 进行能量优化，分四个步骤：

① 固定所有重原子，用共轭梯度法对氢原子进行 100 步优化；

② 取消对重原子的固定，设定非共价键作用的作用距离为 0.8 nm，并取与距离相关的介电常数，用最速下降法进行 100 步优化；

③ 同上条件，用共轭梯度法进行 2 000 步优化；

④ 在蛋白质周围加三层水分子，同上条件对蛋白质-溶液体系进行 2 000 步共轭梯度优化。

（3）确定评估结构的上下界

一个结构的评估分数是它的每一残基与局部环境的倾向性分数总和。选取 45 个序列同源性小于 25%、分辨率大于0.2 nm、R 因子小于 0.2 的单域蛋白质结构，计算其评估分数s，且在双对数图上用黑圆点表示。如图2.7.14 所示，评估分数与蛋白质的长度l呈线性关系。

$$\ln(s) = -0.811 + 1.046 \times \ln(l), \tag{2.7.15}$$

($\ln(s) = -0.830 + 1.008 \times \ln(l)$，在 Profile-3D 中)

即图 2.7.14 中经过黑圆点的拟合直线。

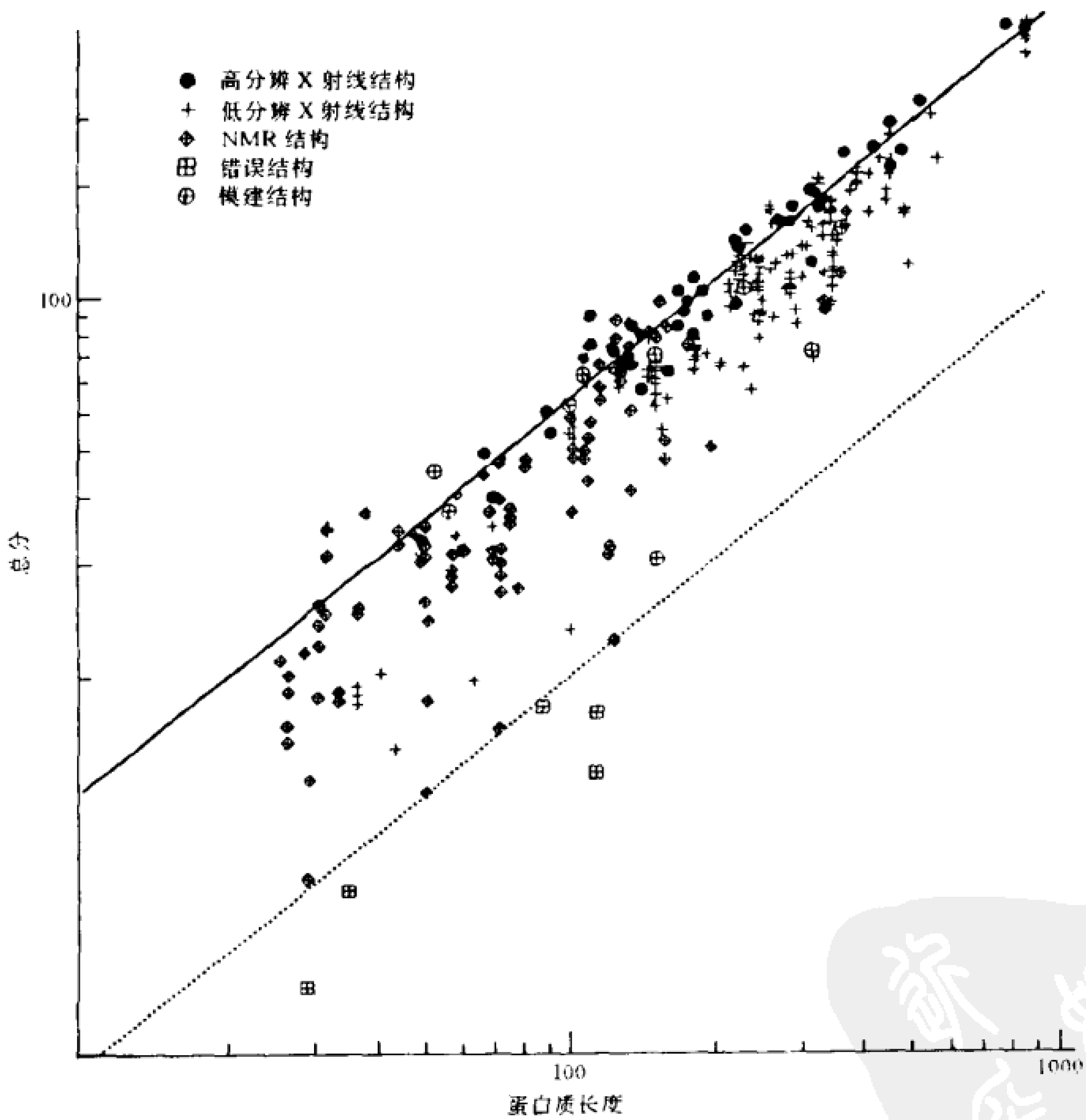


图 2.7.14 评估蛋白质结构-序列兼容性的分数图

横轴是序列长度，标度是双对数尺度。实线由高分辨结构最小二乘法拟合（分辨率至少是 0.2 nm，R 因子小于 0.2），下方的虚线由错误结构拟合。

然后计算 PDB 中所有蛋白质结构与其序列的兼容性的评估分数，这些结构包括分辨率小于 0.2 nm 的结构、NMR 结构以及上述构建的错误蛋白结构。把这些分数全部表示在双对数图上。平行向下移动此拟合直线，直到包含尽可能多的 PDB 结构而不包括错误结构，最后确定的下界直线为

$$\ln(s) = -2.513 + 1.046 \times \ln(l), \quad (2.7.16)$$

$$(\ln(s) = -1.629 + 1.008 \times \ln(l), \text{在 Profile - 3D 中})$$

即图 2.7.14 中的虚线，位于它下面的点均指错误结构，与 Profile - 3D 的判定一致。对于一定长度的蛋白质，由(2.7.15)和(2.7.16)式可以确定评估分数的上界与下界。评估分数处在上界以上的蛋白质结构被认为是完全正确的；若在下界以下，则是完全错误的结构。当评估分数处在上界和下界之间，则蛋白质整体折叠是正确的，但是局部折叠可能有错误。我们采用一个长度为 11 个残基的窗口在蛋白质链上逐步滑动，窗口中残基与

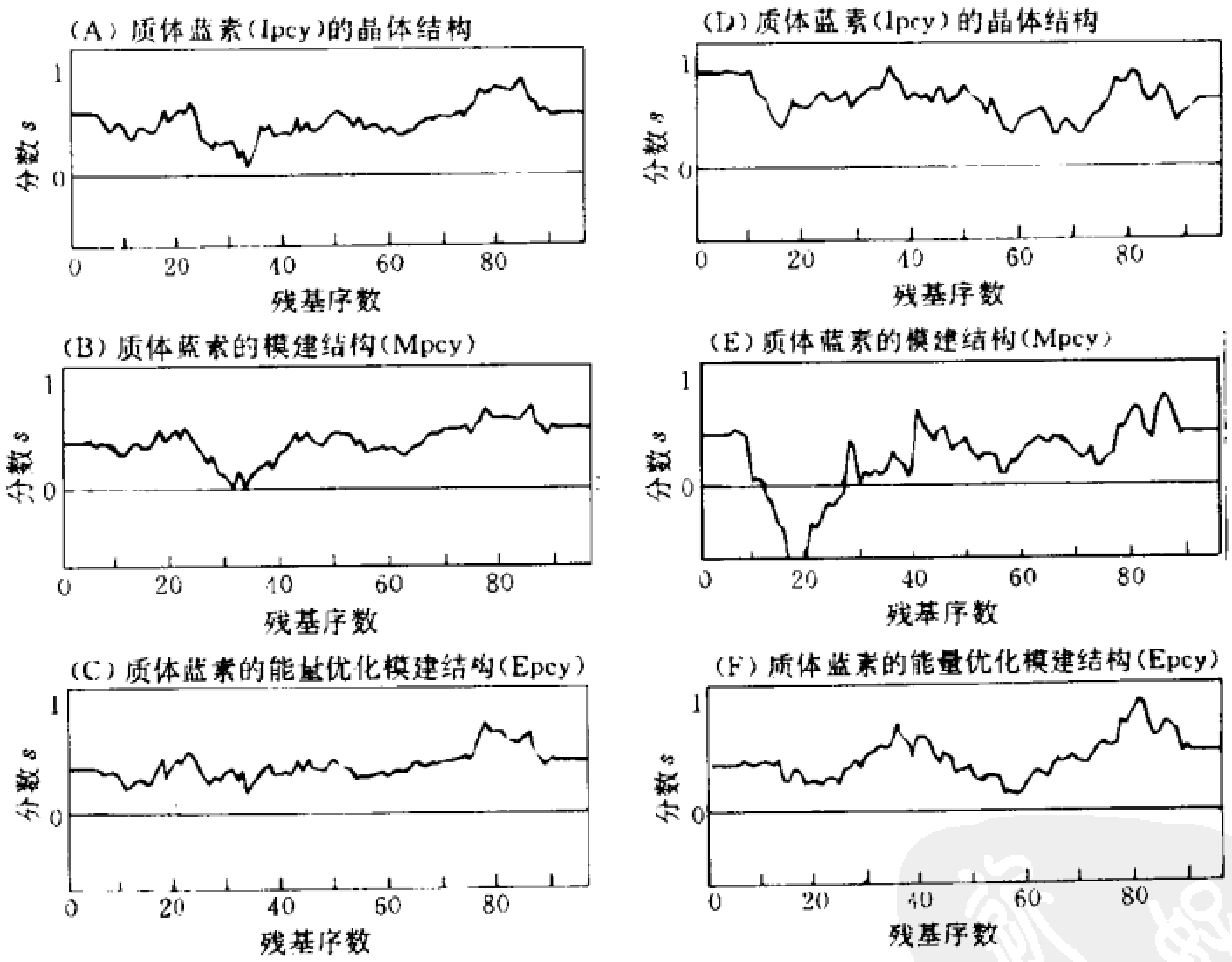


图 2.7.15 用软件 Profile - 3D 和 ASSESSOR 评估结构

(A)、(B)和(C)分别表示 Profile - 3D 评估晶体结构、模建结构和能量优化模建结构的结果；(D)、(E)和(F)表示 ASSESSOR 评估相应结构的结果。

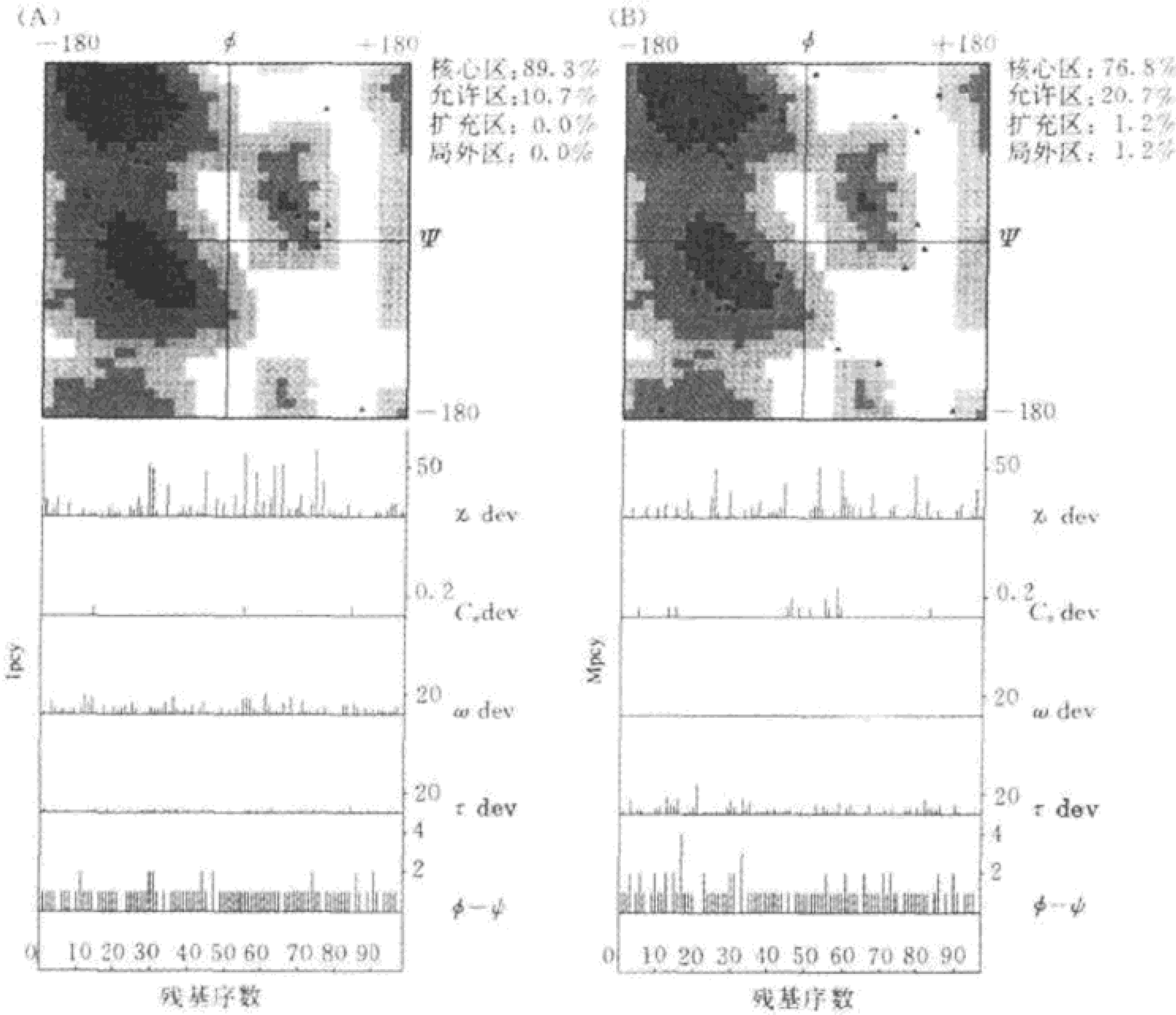


图 2.7.16　对于晶体结构 IPCY(A)和模建结构 MPCY(B)图示在各残基处的各种立体化学参数值

注：这两个蛋白质中每个残基的主链两面角对 $\phi-\psi$ 都画在 Ramachandran 图中，其中甘氨酸 Gly 和脯氨酸 Pro 残基用三角形表示，而其它残基用方块表示。黑色区域表示 $\phi-\psi$ 出现频率最高的核区 CORE，深灰区域 ALLOWED 中的出现频率次之，ALLOWED域向外扩充 20°的区域 GENEROUS 用浅灰色表示，剩余区域是出现频率极低的局外区 OUTSIDE。每个蛋白的 $\phi-\psi$(扣除 Gly 和 Pro)在各区的出现频率在图的右上角标示。在图下方，各水平线表示各种参数的期望值，χ 表示侧链的第一两面角，C^a 表示相邻 a 碳原子的距离，ω 表示肽键角，τ 表示虚拟 C^a—N—C—C^β 两面角。在各残基处每一个参数与其期望值的偏差用垂直线表示。角度单位是度，C^a 距离单位是10^{-1}nm。底图是 $\phi-\psi$ 分布，分布值1(图中省略)表示CORE，2 表示 ALLOWED，3(图中省略)表示GENEROUS 和4 表示OUTSIDE。dev 是偏差的简略符号。

环境的倾向性分数的平均值作为中间残基的分数，沿蛋白质链形成一条评估曲线。利用该曲线可以评估局部结构的正确与否，落在零线以下的区域被认为是错误可能性比较大的区域。例如图 2.7.15(E)中 Mpcy 的剖面图中残基 20 位附近有一个明显的负峰，这是由于在 15 位、22 位和 35

位存在3个不正确的正 ϕ 角的Proline而引起的。

2. 立体化学参数的检验

Morris的研究表明蛋白质中的一些立体化学参数的分布具有一定的规律性，并且随着分辨率的提高，这些参数的分布越来越集中。例如，主链二面角 $\phi-\psi$ 分布在二面角空间的几个特定区域，如图2.7.16所示的Ramachandran图，分布概率最高为核心区(CORE，图中的黑色阴影区)，分布概率次之为允许区(ALLOWED，图中的深灰色阴影区)，允许区向外扩充20°为扩充区(GENEROUS，图中的浅灰色阴影区)，剩余区域为概率极低的局外区(OUTSIDE)。表2.7.1列出了Morris给出的理想的立体化学参数。它们对于模建结构来讲相对有些苛刻，并不能作为结构评估的判据，但可作为评估的参考。我们利用这些参数编制了PROSCAN软件，给出详细的立体化学参数的图象和数值信息。

表2.7.1　高分辨结构的立体化学参数期望值

<table>
<tr><th colspan="2">立体化学参数</th><th>期望值</th><th colspan="2">立体化学参数</th><th>期望值</th></tr>
<tr><td colspan="2">核内 $\phi-\psi$ 百分数</td><td>>90%</td><td colspan="2">ϕ(脯氨酸)</td><td>−65.4°±11.2°</td></tr>
<tr><td colspan="2">χ_1(正邻位)</td><td>−66.7°±15.0°</td><td rowspan="2">螺旋角</td><td>ϕ</td><td>−65.3°±11.9°</td></tr>
<tr><td colspan="2">χ_1(负邻位)</td><td>+64.1°±15.7°</td><td>ψ</td><td>−39.4°±11.3°</td></tr>
<tr><td colspan="2">χ_1(反式)</td><td>183.6°±16.8°</td><td colspan="2" rowspan="2">氢键键能</td><td rowspan="2">−8.4±0.75 kJ/mol</td></tr>
<tr><td colspan="2">χ_1 集合的标准偏差</td><td>15.7°</td></tr>
<tr><td rowspan="2">χ_3(二硫键)</td><td>左联</td><td>−85.8°±10.7°</td><td colspan="2">反式 ω</td><td>179.6°±4.7°</td></tr>
<tr><td>右联</td><td>96.8°±14.8°</td><td colspan="2">$C^{\alpha}-N-C-C^{\beta}$ 二面角</td><td>33.9°±3.5°</td></tr>
</table>

3. 小结

ASSESSOR用C语言编制，在SGI-R3000工作站上检验一个蛋白质需1分钟。

(1) 已知结构蛋白质的评估

从PDB中分别选取大于0.2 nm的低分辨率结构、小于0.2 nm的高分辨率结构和NMR结构若干，尽量使它们不具有同源性。用ASSESSOR和Profile-3D软件分别对它们进行评估，结果列表2.7.2，两者的基本趋势是一致的，但前者比后者更与晶体结构分辨率的高低趋势相符。例如，分辨率为0.155 nm的1paz及0.165 nm的9pap，ASSESSOR的评

估分数在上界以上，而 Profile - 3D 的评估分数则在上界与下界之间；对于分辨率为 0.196 nm 的 1lzt，则正好相反，Profile - 3D 的评估分数在上界以上，而 ASSESSOR 评估分数在上界和下界之间。这种情况的出现与打分机制相关。

表 2.7.2 用软件 ASSESSOR 和 Profile - 3D 计算 PDB 结构与几个错误结构的比较

PDB 标码	长度	分辨率 (nm)	ASSESSOR 分数		Profile - 3D 分数	
1ycc	108	0.123	74.1	(59.5～10.9)	49.2	(48.7～21.9)
3b5c	85	0.15	49.5	(46.3～8.4)	38.1	(38.3～17.2)
1ccr	111	0.15	74.2	(61.3～11.2)	48.4	(50.1～22.5)
2hbg	147	0.15	98.6	(82.2～15.0)	76.8	(66.5～29.9)
1apz	120	0.155	70.8	(66.5～12.1)	43.0	(54.2～24.3)
2ptn	223	0.155	145.3	(127.1～23.2)	108.2	(101.2～45.6)
351c	82	0.16	56.5	(44.6～8.1)	30.5	(36.9～16.6)
1pcy	99	0.16	62.4	(54.4～9.9)	47.1	(44.7～20.1)
9pap	212	0.165	140.5	(120.5～22.0)	93.9	(96.2～43.3)
6xia	387	0.165	218.1	(226.2～41.2)	168.9	(176.5～79.4)
1ofv	169	0.17	96.0	(95.1～17.3)	85.5	(76.5～34.4)
1caa	53	0.18	32.0	(28.3～5.2)	20.7	(23.8～10.7)
2cdv	107	0.18	88.1	(59.8～10.7)	36.5	(48.3～21.7)
1lzt	129	0.196	70.5	(71.7～13.1)	74.5	(58.3～26.2)
5mbn	153	0.20	86.7	(85.7～15.6)	76.8	(69.2～31.2)
5tnc	161	0.20	104.7	(90.4～16.5)	62.3	(72.9～32.8)
1ca2	256	0.20	154.6	(146.8～26.8)	134.0	(116.3～52.4)
1dtx	58	0.22	29.3	(31.1～5.7)	17.9	(26.5～11.9)
1bbc	124	0.22	52.2	(68.8～12.5)	50.5	(56.0～25.2)
3hvp	97	0.28	43.3	(53.2～9.7)	35.8	(43.7～19.7)
3at1B	146	0.28	62.6	(81.6～14.9)	49.5	(66.1～29.7)
1pad	212	0.28	121.7	(120.5～22.0)	97.0	(96.2～43.3)
1gcn	29	0.30	−6.0	(15.0～2.7)	1.1	(12.9～5.8)
3xia	377	0.30	−3.1	(220.1～40.1)	72.7	(171.9～77.3)
5rnt	104	0.32	66.0	(57.2～10.4)	48.1	(46.9～22.1)
1trcA	68	0.36	20.4	(36.7～6.7)	15.1	(30.6～13.8)
1mrt	31	NMR	21.0	(16.1～2.9)	11.3	(13.9～6.2)
1epg	53	NMR	13.2	(28.3～5.2)	20.0	(23.8～10.7)
1hcc	59	NMR	21.5	(31.6～5.8)	21.5	(26.5～11.9)
1c5a	65	NMR	33.8	(35.0～6.4)	18.9	(29.2～13.1)
1acp	77	NMR	16.9	(41.8～7.6)	16.0	(34.7～15.6)
4trx	105	NMR	39.0	(57.8～10.5)	45.2	(47.4～21.3)
2cp1	227	NMR	123.7	(129.5～23.6)	98.9	(103.1～46.4)
Bmcp	113	modeled	7.9	(62.4～11.4)	3.13	(51.0～23.0)
Bhmd	113	modeled	5.5	(62.4～11.4)	17.3	(51.0～23.0)
Bgn5	87	modeled	−0.1	(47.5～8.7)	7.8	(39.2～17.6)
Bhpr	87	modeled	8.2	(47.5～8.7)	6.5	(39.2～17.6)

注：最后 4 个结构是错误结构，其他均为 PDB 结构。圆括号中的数值是公式(2.7.15)和(2.7.16)算得的上界与下界分数。

（2）模建结构的评估

评估方法在蛋白质模建工作中极为重要，好的评估方法能够给出蛋白质折叠是否正确以及局部结构是否错误的信息。分析这些信息，人们能够对模建过程中存在的错误及时修正。

模建的结构分别由我们的软件 PMODELING 和 COMPOSER 获得。ASSESSOR 与 Profile－3D 分别对初始模建结构和经过能量优化后的结构进行评估，分数列于表 2.7.3。比较发现，两者存在一定分歧，经能量优化后，ASSESSOR 对模建结构的评估分数都有明显提高，而 Profile－3D 评估分数提高不大，甚至下降，我们以质体蓝素和牛胰蛋白酶的回顾性模建为例给予详细讨论。

1）质体蓝素的模建评估

表 2.7.3　用软件 ASSESSOR 和 Profile－3D 评估结构的比较

PDR标码	长度	ASSESSOR 分数				Profile－3D 分数			
		上界	下界	初始模建	优化结构	上界	下界	初始模建	优化结构
pcy	97	53.2	9.7	29.6	51.3	43.7	19.7	42.8	42.8
ptn	221	125.9	23.0	81.4	104.0	100.3	45.1	101.0	104.2
mbn	145	81.0	14.8	64.6	69.5	65.6	29.5	64.7	65.3
ccr	104	57.2	10.4	29.7	61.4	46.9	21.1	45.3	44.3
bbc	122	67.6	12.3	38.3	63.4	55.1	24.8	45.4	46.6
dtx	55	29.4	5.4	25.7	27.2	24.7	11.1	15.8	13.2
caa	51	27.2	5.0	23.0	34.7	22.9	10.3	17.8	20.3
lzt	127	70.5	12.9	62.6	69.4	57.4	25.8	70.6	64.6
mpp	347	201.8	36.8	104.5	149.8	158.1	71.1	156.6	166.5
Mp16	148	82.8	15.1	—	20.2	67.0	30.1	—	35.9
Mcdk	303	175.1	31.9	—	70.8	137.9	62.0	—	100.7

注：无数字前缀的 PDB 标码表示由软件 PMODELING 模建的结构；冠以 M 的 PDB 标码表示由软件 COMPOSER 模建的结构。上界兼容分数和下界兼容分数分别用公式（2.7.15）和（2.7.16）计算。

用拟天青蛋白 1paz 和质体蓝素 7pcy 作为模板模建质体蓝素 1pcy，它们与 1pcy 的序列同源性分别为 23%和 55%。在 PMODELING 的模建中，假设 Proline 均按反式构建，模建蛋白用 Mpcy 表示。Profile－3D 的判断认为接近上界，而我们的方法评估分数处在上界和下界的中间部分。从图 2.7.15(E)中 Mpcy 的剖面分析图可以发现在残基 20 位（0 位为 N

末端)附近有一负峰,其主要原因是存在三个不正常的正 ϕ 角的 Proline。统计表明[8],Proline 的 ϕ 角一般处在 $-65.3°\pm11.9°$,而 Mpcy 在 15 位、22 位和 35 位的 ϕ 分别为 110.0°、32.8°和 86.5°。经过能量优化后,15 位、22 位和 35 位的新 ϕ 角分别为 $-89.0°$、$-59.7°$和 $-87.9°$。我们的方法重新对优化后的结构(Epcy)评估,发现其分数有很大提高,从原来的 29.6 提高到 51.3,接近上界分数 53.2。而且从剖面分析图可以看出分数曲线处处在零线以上,如图 2.7.15(F)中 Epcy 剖面图所示。而 Profile-3D 对能量优化前后模建蛋白的评估分数均为 42.8。图 2.7.15(B)中的Mpcy和图 2.7.15(C)中 Epcyd 的剖面图表明 Profile-3D 对明显不符合立体化学参数约束的局部结构错误并不灵敏。其 Mpcy 在 35 位附近出现的一个低峰在晶体结构中也有(图 2.7.15(A)),并非是对 35 位出现的正 ϕ 角的正确反映。用 PROSCAN 对 Mpcy 进行分析也发现了同样的错误,同晶体结构相比(图 2.7.16(A)),$\phi-\psi$ 角处在 Ramachandran 图中核心部位的比例下降了 12.4%,而且相邻 C^{α} 原子间距和残基的虚二面角 C^{α}—N—C—C^{β} 的偏差比较大,如图 2.7.16(B)所示。

2) 牛胰蛋白酶的模建评估

我们用牛胰凝乳蛋白酶 A 链(4cha)、猪弹性蛋白酶(3est)和老鼠巨细胞蛋白酶 A 链(3rp2)模建牛胰蛋白酶 2ptn,它们与目标蛋白的序列同源性分别为 45%、39%和 33%。在 PMODELING 模建中也假设 Proline 均按反式模建。模建结构(Mptn)和晶体结构的 C^{α} 原子的根均方差为 0.937。Profile-3D 判断 Mptn 为评估分数处于上界的理想蛋白,而我们的评估分数处在上下界之间,检测发现 Mptn 并非是没有缺陷的理想蛋白,它存在下列一些结构缺陷:

① 许多相邻 C^{α} 原子的间距偏离 0.38 nm,其中最大的为 0.588 nm,最小的为 0.036 nm。而对应的晶体结构 2ptn 中的所有相邻 C^{α} 原子的间距在 0.37~0.39 nm 之间。

② 其中剖面分析图中一个最大的负峰位置 152 位是正 ϕ 的 Proline。

③ 在 2ptn 中有 6 对二硫键,然而在模建结构中少了两对二硫键。

经过能量优化后,我们对优化结构(Eptn)的评估分数由原来的 81.4 提高到 104.0。分析优化结构的剖面图可以发现,优化前的多个处于零线

以下的区域升到零线以上，只有152位仍是负峰，原因是 Proline 的正 ϕ 角在优化过程中没有得到释放。这说明能量优化并不能解决所有在模建过程中产生的结构歧变(很可能是局部能量极小陷阱现象)。Profile－3D 对优化后结构的评估分数提高不大，由原来的101.0到104.2。从 Profile－3D 对初始结构和优化结构的剖面分析图无法发现这些局部结构错误。由 PROSCAN 的评估图中可以同样发现，Mptn 的 C^{α} 原子的间距偏差比较大，虚二面角的偏差也比较大。

由于 Profile－3D 的方法对局部环境的划分考虑了二级结构、溶剂可及性和侧链被极性原子包埋的面积，这些参数都是描述蛋白质整体折叠性质的，因此它能对蛋白质整体折叠是否正确作出判断，通过我们的方法也证明了这一点。然而，由于在它的结构环境划分中描述局部结构的特征量欠直接和细致，因而对局部结构的歧变不很敏感。我们的方法由于直接考虑了侧链氢键连接以及间接考虑了二硫键这些决定蛋白质局部结构的作用，而且用了多重结构联配，从 $N_{si,j}$ 开始统计，相当于大大扩大了统计样品，因而既能对蛋白质的整体折叠作出判断，又能检测出局部结构出现的错误。

为什么在我们模建的结构中出现正 ϕ 角的 Proline 以及相邻 C^{α} 原子的间距容易偏离 0.38 nm 呢？这是由于上述 PMODELING 模建中假设 Proline 均按反式安装，没有区分 Proline 的顺式和反式情况。实际上，当相邻的 C^{α} 间距小于 0.32 nm 时，Proline 应该按顺式构建，反之按反式构建。按照这种区分，PMODELING 模建的结构 Mpcy 的评估分数为47.2，接近上界值53.2，并且剖面评估的分数处处在零线以上。相邻 C^{α} 原子的间距容易偏离 0.38 nm 可能由于在模建 C^{α} 原子时采用了模板结构加权平均的方法，有时很难保证 C^{α} 原子间距在 0.38 nm 附近。因此在构建 C^{α} 原子时考虑到这个约束条件，引入正则化的方法或能量优化的方法可能会消除这种缺陷。由此可见，我们的评估能够对模建过程作出指导作用。同时，我们的结果也说明不能仅用原子之间的根均方差来评价模建结构的优劣。

参考文献

[1] Ding Dafu, Qian Jiang, Feng Zukang. A differential geometric treatment of Protein Struc-

ture Comparison. *Bull Math Biol*, 1994, **56**(5): 923

[2] 张保红,丁达夫,蛋白质结构成对比较的新方法. 生物物理学报, 1993, **9**(3): 353

[3] 钱江,陈洁,丁达夫. 自然氨基酸的溶剂可及性标准. 生物物理学报, 1993, **9**(2): 314

[4] 陈洁,丁达夫. 蛋白质侧链的预测,生物物理学报, 1993, **9**(4): 651

[5] Ding Dafu, Tang Hai - xu. An improvement of Myers - Miller's algorithm for sequence alignment. *J Biomath*, 1994, **9**(2): 1

[6] 汤海旭,丁达夫. 用于蛋白质同源模建及三级结构预测的结构比较方法,生物物理学报, 1995. **11**(1): 60

[7] 冯祖康,丁达夫. 用于蛋白质分子设计的环区模建方法. 生物化学与生物物理学报, 1995, **27**(2): 173

[8] 张保红,丁达夫,汤海旭. 一种快速和优质的蛋白质结构刚体叠加方法. 生物物理学报, 1995, 11(4): 575

[9] 丁达夫,汤海旭,张保红. 基于结构比较的蛋白质模建系统及其评估. I,主链的模建. 生物物理学报, 1995, **11**(3): 416

[10] 汤海旭,丁达夫. 基于结构比较的蛋白质模建系统及其评估. II. 侧链的安装. 生物物理学报, 1996, 12(1): 125

[11] 张保红,丁达夫. 基于结构比较的蛋白质模建系统及其评估. III. 灵敏的蛋白质结构的评估方法,生物化学与生物物理学报, 1996, 28(4): 335

[12] Zheng Q, *et al.* Loop closure via bond scaling and relaxation. *J Comput Chem*, 1993, **14**: 556

[13] Feng Zukang, Ding Dafu. Modeling Method of loop for protein engineering. *J Biomath*, 1995, **10**(1): 1

[14] 汤海旭,丁达夫. 基因组序列的大尺度同源性分析. 生物化学与生物物理学报, 1996, **28**(6): 686

[15] Taylor W R, Orengo C A. Protein structure alignment. *J Mol Biol*, 1989, **208**: 1

[16] Sali A, Blundell T L. Definition of general topological equivalence in Protein structures, *J Mol Biol*, 1990, **212**: 403

[17] Fano R. *Transmission of Information*, New York: Wiley, 1961

[18] Pondes J W, Rivhards F M. Tertiary templates for proteins. *J Mol Biol*, 1987, **193**: 775

[19] Dunbrack R L, Karplus M. Backbone - dependent rotamer library for proteins. Application to sidechain prediction. *J Mol Biol*, 1993, **230**: 543

[20] Desmet J, *et al.* The dead - end elimination theorem and its use in protein side - chain positioning. *Nature*, 1992, **356**: 539

[21] Desmet J, *et al. The "dead - end elimination" theorem. A new approach to the side - chain packing problem. In: The protein folding problem and tertiary structure prediction.* Merz. K M *et al.* eds. *Boston Birkhauses* 1994, 307

[22] Novotny J, *et al.* An analysis of incorrectly folding models. *J Mol Biol* 1984, **177**: 787

[23] Baumann G, *et al.* Charge grouping approaches to calculation of electrostatic forces in molecules dynamics of proteins. *J Biomol Struct Dyn*, 1991, **9**: 205

[24] Holm L, Sander C. Evalution of protein models by atomic solution preference. *J Mol Biol*, 1992, **225**: 93

[25] Topham C M, *et al.* An assessment of COMPOSER: a rule - based approach to modeling protein structure. *Biochem Soc Symp*, 1991, **51**: 1

[26] Bowie J U, *et al.* A method to identify protein sequences that fold into a known three-dimensional structure, *Science*. 1991, **253**: 164
[27] Luthy R, *et al.* Assessment of protein model with three-dimensional profiles. *Nature*, 1992, **356**: 83
[28] Morris A L, *et al.* Stereochemical quality of protein structure coordinates. *Proteins*, 1992, **12**: 345
[29] Laskowski P A, *et al.* PROCHECK-a program to check the stereochemical quality of protein structures. *J Appl Cryst*, 1993, **26**: 283

(丁达夫 汤海旭 张保红 陈 洁)

§2.8 蛋白质结构层次中的超二级结构

近年来关于蛋白质超二级结构的研究成为国际上一个热点课题,国内也开始出现有关超二级结构的研究论文。蛋白质超二级结构是两个或几个规则的二级结构单元的进一步组合,或看成是二级结构的局域折叠。

2.8.1 连接多肽区

在蛋白质结构中,α螺旋和β折叠等二级结构单元进一步组合,折叠成三维空间结构。在这种肽链的折叠过程中,α螺旋和β折叠被各种长度的"走向"不规则的连接多肽(loop)连接而成结构模体。高分辨率的X衍射晶体蛋白的构象分析表明,α螺旋和β折叠的连接短肽的构象类型是有限的[1~3],这些连接短肽可以反向改变链走向(180°),或使链密集,或改变90°方向。有些连接肽则使链微微弯曲,或发生扭结,它们在蛋白质肽链折叠过程中具有明确的结构作用。

这些连接多肽区的主链上羰基氧和氨基氮间一般不形成氢键,当其暴露在水溶剂中时,与水分子相互间形成氢键。相当多的连接多肽位于蛋白质分子表面。位于蛋白质分子表面的这些连接多肽暴露在溶剂中,肽链上的残基多是带电残基,或极性残基,或甘氨酸残基。不同种属的同源蛋白质氨基酸序列比较发现,一些残基的插入、替换、删除多发生在连接多肽区,这说明在蛋白质进化演变过程中,蛋白质分子内核比分子表面的连接多肽区更稳定,即连接多肽区柔性易变[4]。

连接多肽除了作为二级结构元件的连接功能外,有些连接多肽还常参与形成一些(如DNA、金属原子等)结合位点和酶的活性位点。

2.8.2 蛋白质中的简单超二级结构

在蛋白质结构中，常常发现两个或几个二级结构元件进一步组合成有特殊几何排列的结构块，这些结构块被称为超二级结构。图2.8.1所示是由两个 α 螺旋与连接多肽组成的最简单的具备特殊功能的超二级结构。图2.8.1(A)为 DNA 结合超二级结构，(B)为钙原子结合超二级结构，两个 α 螺旋间的连接多肽区结合钙原子。

α - loop - α 型超二级结构提供了在合适的位置结合或释放钙配位体的结构骨架。这是在蛋白质结构中识别的第一个功能超二级结构，该超二级结构包含两个 α 螺旋及 12 个残基长的连接多肽区，其中 5 个残基是钙配位体，其侧链包含一个氧原子。

由 β 折叠链形成的简单的超二级结构，是两条反平行 β 折叠链由连接多肽连接而成，这一超二级结构称为 β 发夹，常发生在反平行 β 折叠结构中，作为隔离 β 带或更复杂的 β 折叠片的一部分[5]。β 折叠链之间连接肽长度普遍为 2～5 个残基，该类超二级结构无特殊生物功能。

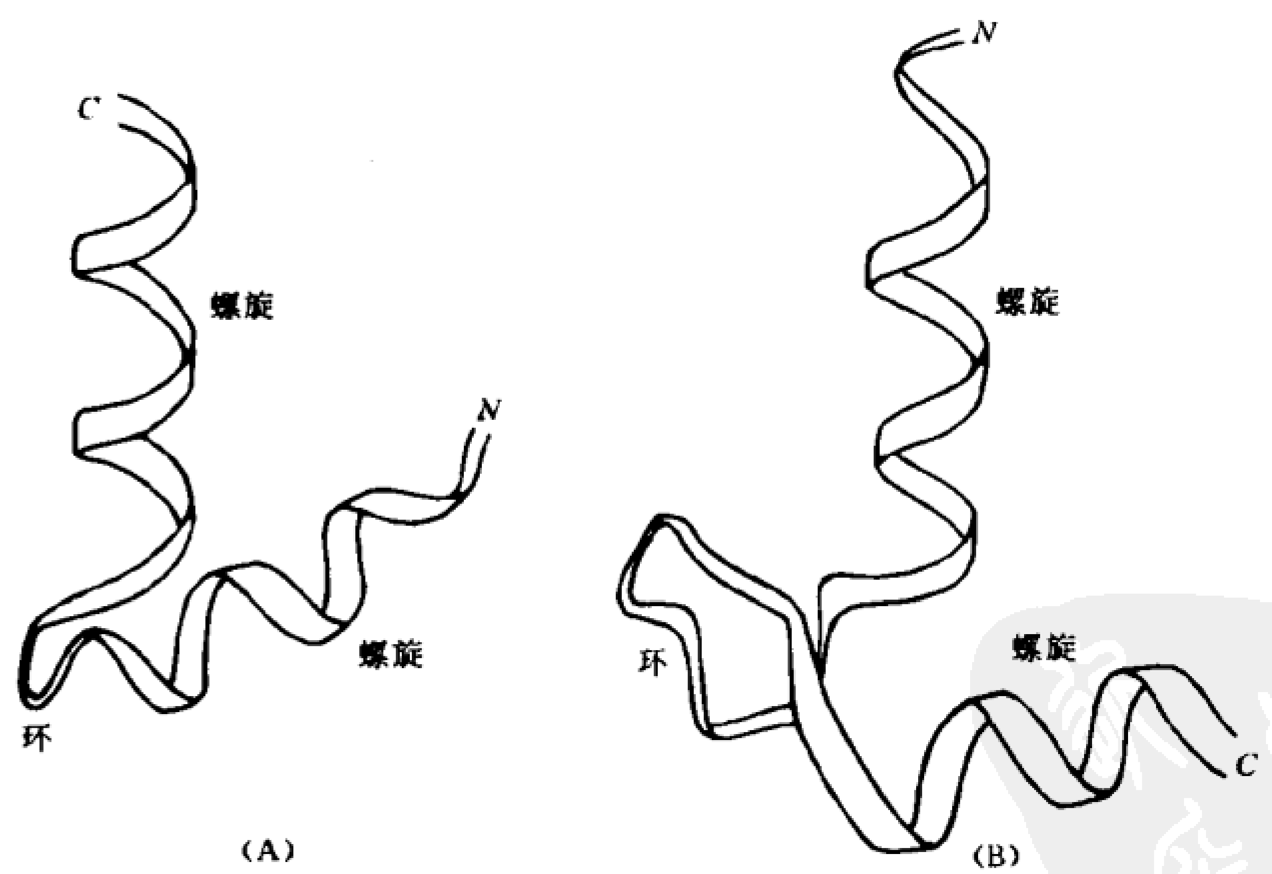


图 2.8.1 两条 α 螺旋由一条短连接肽连接形成一个有特定几何排布的“螺旋 - 连接肽 - 螺旋”超二级结构

(A)DNA 结合超二级结构；(B)钙结合超二级结构。(引自 Carl Branden 等)

2.8.3 在蛋白质结构中频繁发生的超二级结构类型

我们从精选的240个蛋白质的高精度结构中[3]研究发现，超二级结构的构象类型是相当有限的，共4种类型，11种构象形式[6]，并非是可能构象的随机组合。它们与连接肽所连接二级结构单元的种类有关，与连接肽长度、连接肽残基构象有关。这4种超二级结构类型是α－α拐角，α发夹，β发夹，拱形结构[6,7]。在α－α拐角中，α螺旋经过连接肽转了90°角，或者表述为连接肽连接两个不在一个平面的两条α螺旋，其矢量夹角为90°。α发夹是两条平行的α螺旋被连接肽所连接。β发夹是两条反平行β折叠链被连接肽连接。拱形超二级结构则为连接肽连接两个不同类型的不在同一平面的二级结构。图2.8.2是频繁发生的4种类型简单的超二级结构的拓扑结构图[6]，图2.8.3是它们所包含的11种构象和氢键模式图[3]。

这些超二级结构中连接多肽的亲水性、疏水性都有其特定的规律，即使出现在非同源蛋白质结构中，超二级结构的核团也总是以相同方式充填。它们具有确定的序列模式特征，带电残基、极性残基、以及甘氨酸的出现都有其一定规律[6,8]。

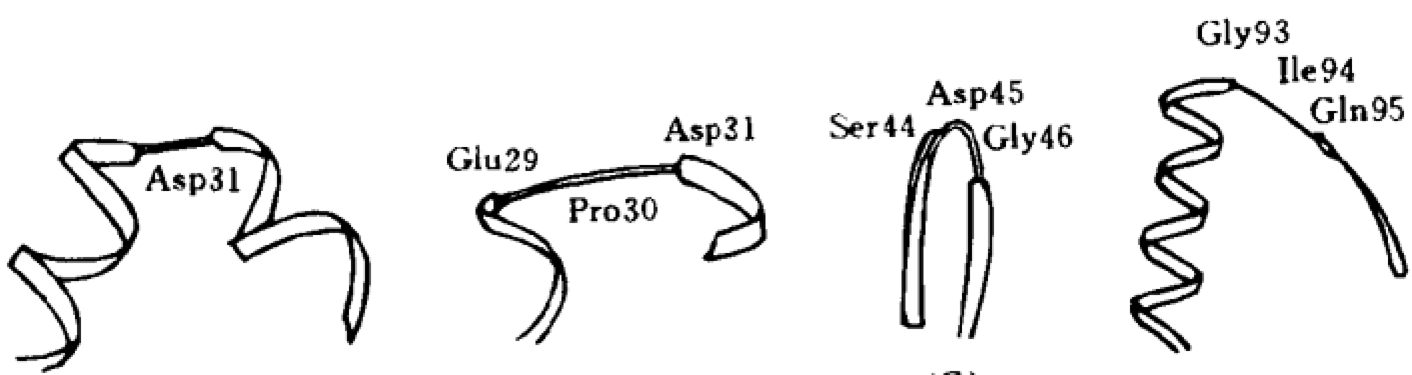


图2.8.2 四种类型超二级结构拓扑图

(A)α－α拐角；(B)α发夹；(C)β发夹；(D)拱形结构。

2.8.4 简单的超二级结构进一步组合形成复杂的超二级结构

一些简单的超二级结构可以进一步组合形成复杂的超二级结构。例如两个紧邻的β发夹超二级结构可以进一步组合，其可能的组合方式有12种。它们能够组合成更复杂的超二级结构。四个紧邻的反平行β折叠链常组成的模式，称为希腊钥匙。希腊钥匙不参与任何专一功能，但在蛋白质结构中常出现。

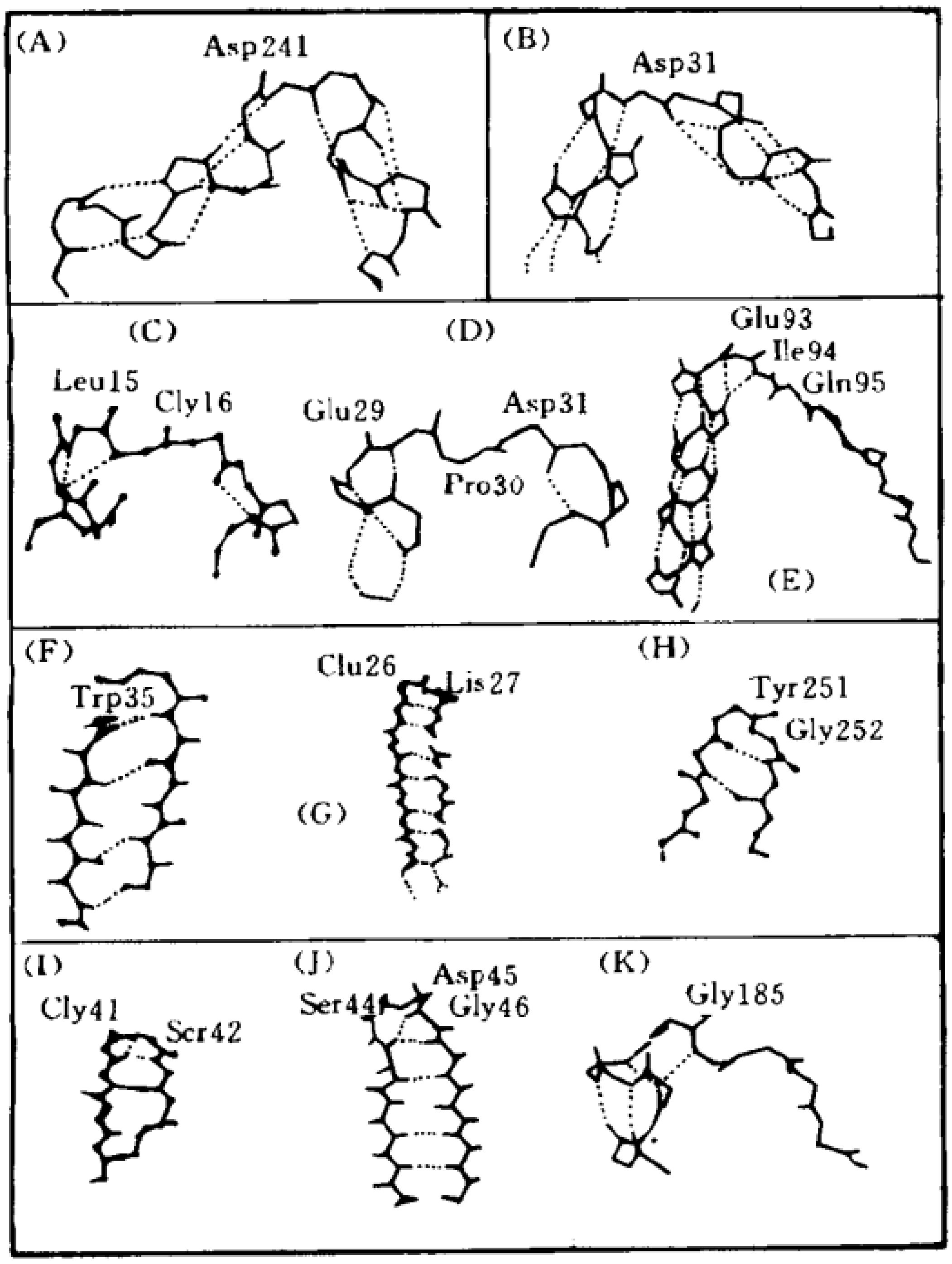


图 2.8.3 11 种构象类型的超二级结构构象和氢键模式图

(A),(B)$\alpha-\alpha$ 拐角;(C),(D)α 发夹;

(F),(G),(H),(I),(J)β 发夹;(E),(K)拱形结构。

在有的蛋白质结构中,β 折叠链的交叉连接回折通过 α 螺旋实现。这时肽链可能通过连接肽回折两次,最终超二级结构组成是 β - loop - α - loop - β。简写成 $\beta-\alpha-\beta$ 超二级结构。例如在磷酸丙糖异构酶中,包含多个 $\beta-\alpha-\beta$ 结构。几乎在每一个有平行 β 折叠片的蛋白质中都发现这一结构,因此 $\beta-\alpha-\beta$ 超二级结构由两个相互平行的 β 折叠链、α 螺旋和两个连接肽组成。连接肽长度可从一两个残基到十几个残基之间变化。在这个超二级结构中的两个连接肽功能不同,连接 β 折叠链的 C 端和 α 螺旋的 N 端的连接肽常与形成功能结构位点和活性位点相关。$\beta-\alpha-\beta$ 超二级结构有手性,在已知蛋白质中发生的 $\beta-\alpha-\beta$ 超二级结构的手性基

本相同，为右手性。当几个这样的超二级结构连接形成结构域时，该手性对蛋白质的功能有重要影响。另一个连接肽还未发现与其活性位点相关。

2.8.5 超二级结构和二级结构组合构成结构域

超二级结构中借助于侧链的堆积，相邻的 α 螺旋和 β 折叠相互更紧密靠近。几个超二级结构通常会结合起来形成紧密的球形结构，这一结构称为结构域。因此结构域可能通过两种途径形成，或是由超二级结构结合排列而成，或是由肽链折叠而成[9]。结构域定义为能够独立折叠成稳定的三级结构的多肽链的一部分。结构域是蛋白质的一个功能单位，蛋白质的不同结构域常常参与不同的功能。例如 λ 阻遏蛋白在多肽链的 N 端有一个结构域，起结合 DNA 的作用，而 C 端的结构域则起到将两个多肽链一起结合到二聚阻遏蛋白分子中的作用。

蛋白质分子中，可以包含单个结构域或多个结构域，有时多达十多个。结构域的内核是由 α－loop－α、α－loop－β、β－loop－α、β－loop－β 等超二级结构构成。结构域可以分成三种结构类型：即 α 结构，β 结构，以及由 β－α－β 组合成的 α/β 结构和 α+β 结构。

蛋白质折叠的中心法则讲到，一级结构确定三级结构，即专一的蛋白质三级结构是由具有独特的氨基酸序列的多肽链自发折叠形成的。然而，近几年来，一组被称为伴娘(chaperones)的蛋白质被发现，它们在多肽链折叠过程中起着催化促进作用。不同的蛋白质三级结构形成需要不同的伴娘。

由此我们可以看到，蛋白质超二级结构是蛋白质空间结构中极重要的一个结构层次，它的研究与蛋白质结构比较模型及空间结构预测、蛋白质肽链折叠的研究密切相关。现已证明超二级结构的构象类型是有限的。对广泛存在的一类超二级结构——β 发夹结构的结构模型已开展了相当深入的研究，现已能对其数学模型化[9]，进行定量描述[10]。这一研究表明，与其他学科和学科分支一样，结构分子生物学在数学模型描述方面前景也是广阔的。最近，我们又对超二级结构进行了预测研究，取得了较好的进展。此项预测比二级结构预测有更高的精度(75%～80%)，这可能是由于两段二级结构片段及其连接肽一起考虑的缘故，也可能是由于超二级结构有确定的构象模式、序列模式、疏水性特征等原因。

参考文献

[1] Efimov A V. *Curreet Opinion in Structure Biology*,1993,**3**:379

[2] Thoynton J M,Sibanda B L,Edwrlo M S, *et al*. *Bioessays*,1988,**8**:63

[3] 孙之荣,Blundell T L. 生物物理学报,1994,**2**:289

[4] Topham C M,Macleod A,Eisenmenger F, *et al*. *J M Biol*,1992,**212**:403

[5] Sibanda B L,Thoynton J M. *Nature*,1985,**316**:170

[6] 孙之荣. 科学通报,1994,**39**(24):2260

[7] Sali A,Blundell T L. *J Mol Biol*,1990,**212**:403

[8] Hubbard T J P,Blundell T L. *Protein Engin*,1987,**1**:159

[9] Janin J,Chothia C. *Methods Enzymol*,1985,**115**:420

[10] Sibanda B L,Thoynton J M. *J Mol Biol*, 1993,**229**:428

(孙之荣)

第3章 基因序列

§3.1 DNA序列的几何学理论及其生物学意义

当前,DNA测序工作正在世界各地的生化实验室里夜以继日地进行,所发表的序列量正以惊人的速度增长。美国GenBank每3个月就要更新存储DNA序列的光盘旧版本。亿万年来生物进化的历史写在这本浩繁巨大的"天书"之中。面对这浩如烟海的数据,如何从中揭示生命的奥秘,探索大自然的规律,历史性地摆在现代人的面前。这是一个严峻的挑战。

DNA序列实际上是一种用4种字母表达的"语言",只是其"词法"和"语法"规则在目前基本上仍不清楚。人类的语言有文字、声音两种基本表现形式,此外还有手语、旗语甚至图画语等特殊表达形式。同样,DNA序列作为一种语言,其表达形式也不是唯一的。传统上,DNA序列是用4种字母符号表达的一维序列。这是一种抽象形式,适合于存储、印刷和代数算法的处理,包括比较、排列和查找特殊序列等。近年来,立足于国内,我们开拓了DNA序列几何表示形式的研究,将DNA序列用三维空间曲线表示。几何形式虽然与符号形式完全等价,但显示了DNA序列的新特征。两种形式各有其特点,相互补充。通过几何形式,我们开始将过去认为与DNA序列风牛马不相及的数学工具,诸如群论、计算几何、微分几何、频谱分析、数字滤波理论等,引入到DNA序列的分析中来。这一新兴研究领域已经展现了广阔的发展前景,为解读DNA序列信息提供了崭新的手段。

3.1.1 DNA序列中的三种分布

我们的研究始于对4种碱基对称性的观察。图3.1.1显示了腺嘌呤A,鸟嘌呤G,胞嘧啶C和胸腺嘧啶T的化学结构。本文按国际生物化学联盟术语委员会(NC-IUB)1984年的命名来表示碱基及其集合[1]。按双

环或单环结构来划分，碱基可分为两类，即嘌呤与嘧啶。可表示为

$$\text{碱基}\begin{cases}\text{嘌呤}, R=A \text{ 或 } G, \\ \text{嘧啶}, Y=C \text{ 或 } T。\end{cases}$$

同样，按环中对应位置上是否存在氨基或酮基来划分，碱基又可分为另外两类，即

$$\text{碱基}\begin{cases}\text{氨基}, M=A \text{ 或 } C, \\ \text{酮基}, K=G \text{ 或 } T。\end{cases}$$

进而，在双螺旋结构中，按碱基对形成氢键的数目或强弱，碱基又可分为另两类，即

$$\text{碱基}\begin{cases}\text{强氢键}, S=G \text{ 或 } C, \\ \text{弱氢键}, W=A \text{ 或 } T。\end{cases}$$

前两种划分为纯化学的，最后一种划分既包含了化学的也包含了 DNA 双螺旋的结构信息在内。

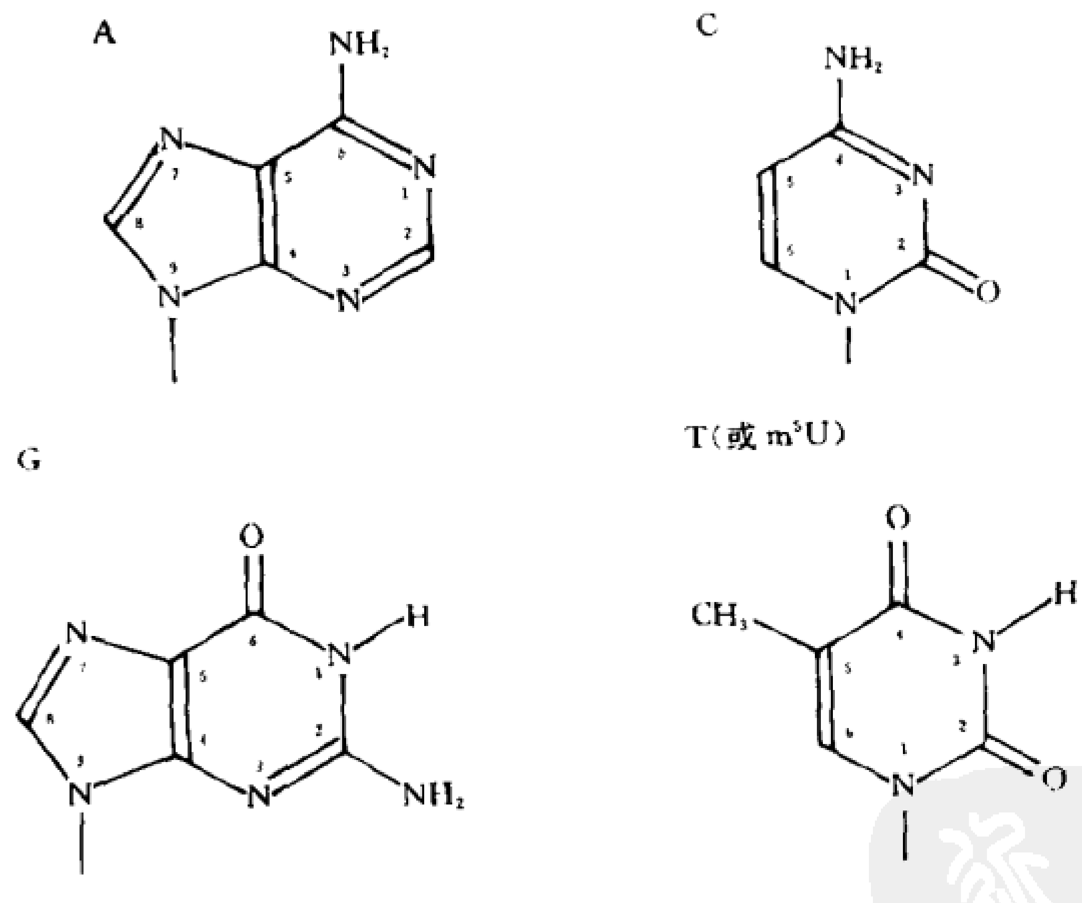


图 3.1.1 四种碱基的化学结构

参照基本粒子理论中的做法，我们利用某种对称的几何图形来表示碱基的上述的对称性。我们首先想到了正多面体。几何定理证明，自然界中只存在五种正多面体，包括正四面体，正六面体等，统称为柏拉图体。现考虑一正六面体，它的六个面按 x,y,z 三个方向可划分为左右，前后和上下三对。不失一般性地，将左右面分别称为嘧啶与嘌呤面；前后面分别

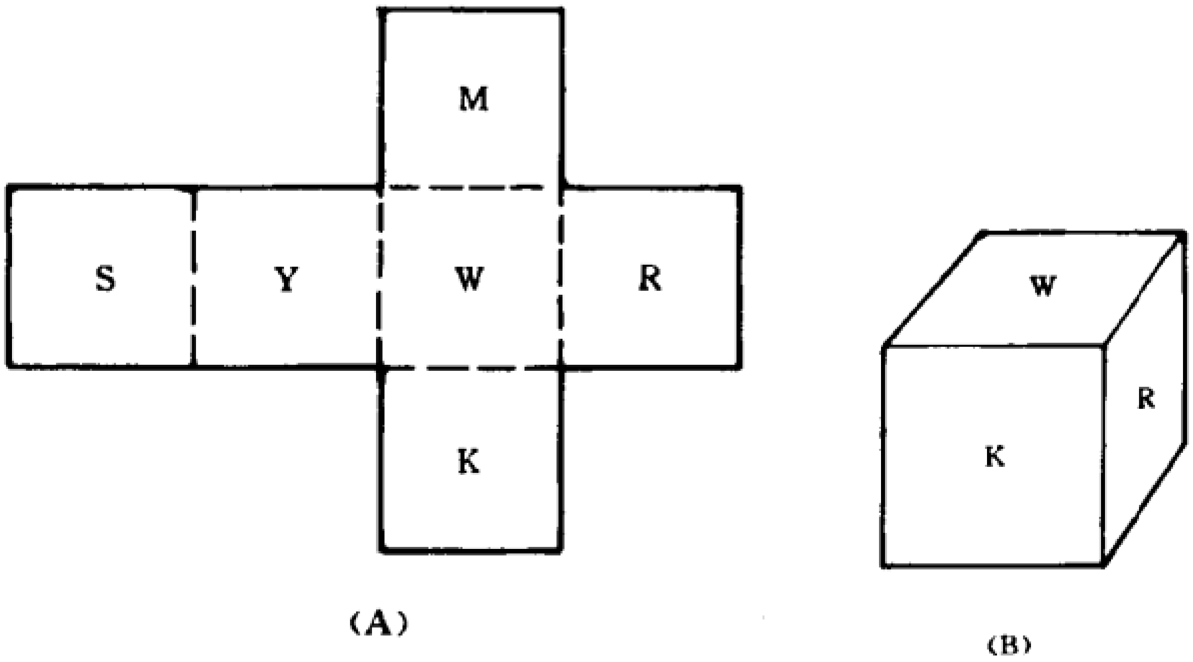


图 3.1.2 用正六面体表示碱基的对称性

为酮基与氨基面；上下面分别为弱氢键与强氢键面，如图 3.1.2 所示。在此表示形式中，嘌呤与嘧啶；氨基与酮基；强氢键与弱氢键两两相对。在六个面中各引一条对角线，使相对面的对角线两两相互垂直，如图 3.1.3 所示。在任意一相对面的两条对角线的四个端点上，分别标上符合该面的碱基符号。例如，不失一般性地，在嘌呤面对角线的两端分别标上 A 和 G；在嘧啶面对角线的两端分别标上 C 和 T，如图 3.1.3 所示。注意，此时前

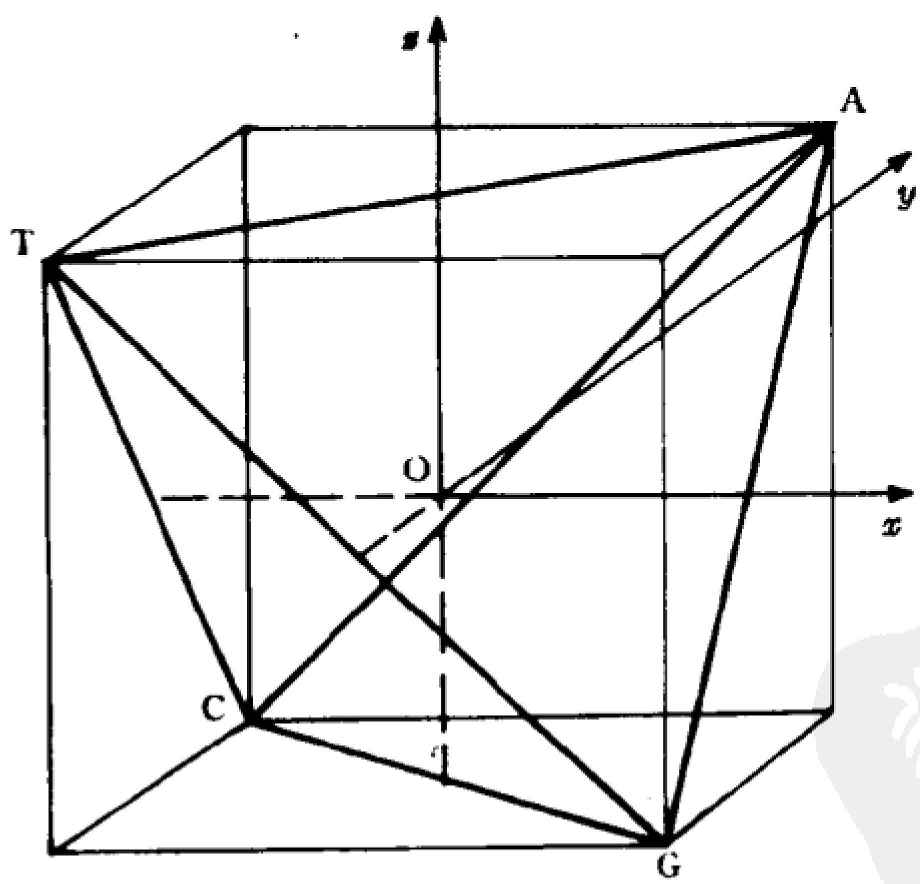


图 3.1.3 用正四面体表示碱基对称性

述碱基对称关系全部自动成立。如上下面分别出现 AT 和 GC，前后面分别出现 GT 和 AC。笔者认为，很难再找到比这更巧妙更合理的几何图形来表达碱基的对称性关系。特别要注意的是，六条对角线刚好是正四面体 ACGT 的六条棱。

以上这种表示形式是笔者在 1991 年首先提出来的[2,3]，并用它来表示 DNA 序列的碱基组成和提出四面体晶格的概念[4]。到 1994 年，又利用这种形式来表示任意长度的 DNA 序列[5]。现将这种序列表示方法简述如下。考察一个长为 L 的单股 DNA 序列，方向（$5'\to3'$ 或 $3'\to5'$）不限。从第一个碱基开始，依次考察此序列，每次只考察一个碱基。当考察到第 n 个碱基时（$n=1,2,\cdots,L$），数一下从 1 到 n 这个子序列中四种碱基各自出现的次数。设 4 种碱基 A、C、G、T 出现的次数分别以 A_n、C_n、G_n、T_n 表示之，这里下标"n"是表明这些整数是从 1 到 n 这个子序列中数出来的，如图 3.1.4 所示。显然，它们都是正整数。根据正四面体的对称性可以证明，在正四面体内存在唯一的一个点 P_n 与这四个正整数对应。点 P_n 构成了四个正整数的一一对应映射。在图 3.1.3 所建立的坐标系之下，点 P_n 的坐标可用四个正整数表达[5]。

$$
\begin{aligned}
&x_n=2(A_n+G_n)-n,\\
&y_n=2(A_n+C_n)-n,\\
&z_n=2(A_n+T_n)-n,\\
&x_n,y_n,z_n\in[-n,n],\ n=1,2,\cdots,L,
\end{aligned}
\tag{3.1.1}
$$

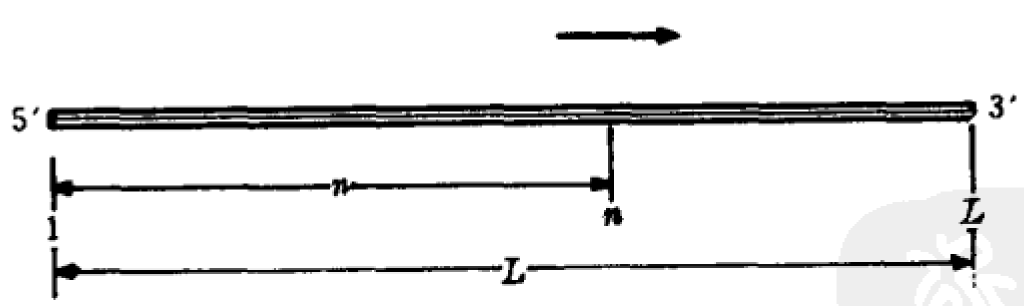


图 3.1.4 DNA 序列示意图

其中 x_n，y_n 和 z_n 为点 P_n 的三个坐标分量。当 n 从 1 跑到 L 时，我们依次得到 $P_1,P_2,\cdots,P_L$ 共 L 个点。将相邻两点用适当的曲线连接所得到的整条曲线，就称为表示此 DNA 序列的 Z 曲线[5]。可以证明，Z 曲线与所表示的 DNA 序列是一一对应的。 即，给定一 DNA 序列，存在唯一的一条 Z 曲

线与之对应;反之,给定一 Z 曲线,可找到唯一的一个 DNA 序列与之对应。换言之,Z 曲线包含了 DNA 序列的全部信息。Z 曲线是与符号 DNA 序列等价的另一种表示形式,一种几何形式。本文的核心想法就是通过对 Z 曲线的研究来对 DNA 序列进行研究。

Z 曲线的三个分量(方程(3.1.1))具有明确的生物学意义:x_n 表示嘌呤/嘧啶碱基沿序列的分布。当从 1 到 n 的这个子序列中(图 3.1.4),嘌呤碱基多于嘧啶碱基时,$x_n>0$,否则 $x_n<0$,当两者相等时,$x_n=0$。同样,y_n 表示氨基/酮基碱基沿序列的分布。当在子序列中氨基碱基多于酮基碱基时,$y_n>0$,否则,$y_n<0$,当两者相等时,$y_n=0$。z_n 表示强/弱氢键碱基沿序列的分布。当弱氢键碱基多于强氢键碱基时,$z_n>0$,否则,$z_n<0$,当两者相等时,$z_n=0$。这三种分布是相互独立的,表现在以下事实上:任何一种分布不能由其他两种分布的线性叠加表示出来。给定的 DNA 序列唯一地决定了这三种分布;三种分布唯一地描述了 DNA 序列。今后,我们对 DNA 序列的研究,就是通过对这三种分布的研究来进行的。从方法学的角度来看,这是 DNA 序列的一种几何学研究途径。

为了给读者一个直观的印象,图 3.1.5 给出了大肠杆菌 *aroP* 基因族序列 Z 曲线的三个分量,即三种分布图。该基因族包含了大肠杆菌 5 个基因 *aroP*, *A*, *aceE*, *aceF* 和 *lpd*, 总长度为9 501 bp,分别编码芳香族氨基酸运输蛋白 *aroP*,蛋白质 A(功能不详)和三种酶,即丙酮酸脱氢酶,二氢硫辛酰胺酰基转移酶和二氢硫辛酰胺脱氢酶。它们位于大肠杆菌遗传联锁图的2.6 min处。这 5 个基因的编码区分别位于此序列的 0 039~1406, 1947~2654, 2870~5527, 5545~7434 和 7759~9183 区间。在图中 X 轴的下方的基因排列图上已分别用阴影区标出。在这些区间之间有三个启动子区,其中 *aceE* 和 *aceF* 基因属于 *ace* 操纵子,共用一个启动子。三个启动子区亦在图中标出。非常令人感兴趣的是,在 5 个编码区,Z 曲线的 z 分量基本上都是单调下降的,而在三个启动子区基本上都是单调上升的。x, y 分量亦有所变化,但不如 z 分量明显。在上升、下降的交界处,Z 曲线均发生了重大的转折,据此有可能用 Z 曲线识别这些位置。由此图可见,用 Z 曲线这种几何方法显示 DNA 序列不仅直观,而且作为一种识别序列中的不同基因和功能区的新方法,展现了广阔的应用前景[6]。

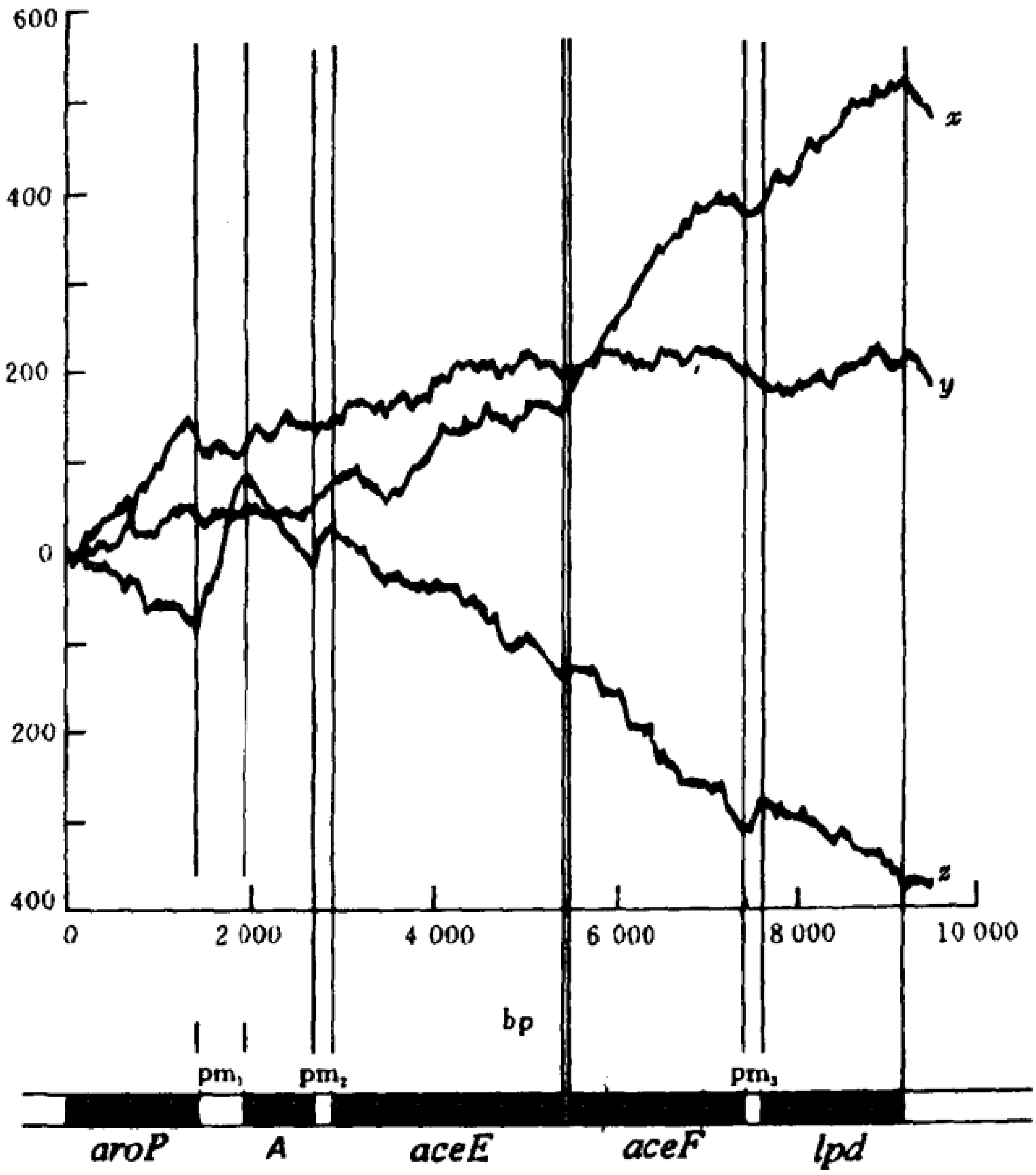


图 3.1.5 大肠杆菌 *aroP* 基因族序列的 Z 曲线的三个分量即三种分布图

此序列包含了五个基因。五个编码区与三个启动子区 pm_1、pm_2、pm_3 的位置已清晰地标出。特别注意 Z 曲线的 z 分量在这 8 个区域及其边界处的行为。

3.1.2 三种分布的对称性研究

既然 DNA 序列由三种分布来完全描述，因此，DNA 序列的对称性将由三种分布的对称性所决定。而三种分布又是基于正四面体的对称性导出的，故正四面体反映了 DNA 序列的对称性。正四面体是一个高度对称的物体，它的对称性表现在绕某些空间轴旋转使它在空间的位置保持不变上。根据群论，共有 12 种旋转满足正四面体的旋转不变性要求。它们全体构成一个有限群，称为正四面体群，简称 T 群。另一方面，4 种物体或符号的置换操作全体形成 4 阶对称群 S_4。其中偶置换全体构成 S_4 群的一个不变子群，称为 4 阶交代群 A_4。当被置换的 4 种物体分别为 A，C，

G,T 四种碱基时,特称此时的 A_4 群为 DNA 群。根据群论,DNA 群与 T 群同构。笔者已利用群论的表示理论建立了 DNA 序列的对称性理论[7],这里着重阐述其生物学意义。由于 DNA 群与 T 群同构,因此 DNA 群中 4 种碱基的变换可通过 T 群中相应的旋转操作来研究。在 T 群的元素中,有一个是绕 z 轴 180°的旋转,将此操作元素记为 R_z。在图3.1.3中将正四面体绕 z 轴旋转 180°,我们发现碱基做了以下的变换:

$$A \to T,\ T \to A,\ G \to C,\ C \to G, \tag{3.1.2}$$

它正是 DNA 群中的互补变换操作元素。此时,

$$\begin{aligned} x_n &\to x'_n = -x_n, \\ y_n &\to y'_n = -y_n, \\ z_n &\to z'_n = z_n, \end{aligned} \tag{3.1.3}$$

将变换(3.1.2)代入方程(3.1.1)亦可直接得到变换(3.1.3)。这一结果的生物学意义为:DNA 双螺旋的双股 DNA 序列具有相同的三种分布,最多相差一个符号。

由于双螺旋的双股是反平行的,因此我们还需要研究将单股 DNA 序列进行反向或反射的操作,即 $5' \to 3'$ 与 $3' \to 5'$ 方向互换,这种操作记为 R_f。已经证明[7],R_f 不改变一个单股序列的三种分布。综上可知,DNA 双螺旋中的双股序列,不论其阅读方向如何,都具有共同的三种分布。

我们发现 Z 曲线的三个分量的平方和 $x_n^2+y_n^2+z_n^2$ 对于任意 n 值是 DNA 群的一个不变量[7],它的生物学意义仍不清楚。同样,DNA 序列的信息熵 S 是 DNA 群的另一个不变量[7]。S 被定义为:

$$S = -\sum_{i=1}^{4} p_i \log_2 p_i, \tag{3.1.4}$$

其中 $p_i(i=1,2,3,4)$ 代表四种碱基在序列中出现的频率。在此基础上,我们引入等熵面的概念,用来研究生物的分子进化[8]。

DNA 序列的对称性表现为,在 DNA 群某元素作用下,DNA 序列的某性质保持不变。根据不同的情况,把对称性分为两类[7]:总体对称性(global symmetry)和局域对称性(local symmetry)。当一个 DNA 序列中的所有碱基都经历了 DNA 群中某元素操作下的变换,而仍保持序列的某性质不变,则称此序列对某变换有总体对称性;反之,若序列中只有个别位置上的碱基经历了某变换,而仍保持了序列的某性质不变,则称此序

列存在局域对称性。前者的例子之一是互补对称性,即在 R_z 作用下 DNA 双螺旋保持不变。后者的例子多见于酶或蛋白质与 DNA 序列的识别方面。下面以第二类限制性内切酶对多重 DNA 序列的识别为例,来讨论局域对称性。

那些能识别特定的 DNA 序列并在特定位置切断序列中的磷酸二酯键的酶称为第二类限制性内切酶[9]。它们是基因工程中的重要工具酶。大多数情况下,一种酶只能识别一个特殊序列,而在少数情况下,识别序列存在局域对称性。例如,AccI 识别序列 5′-GT(A 或 C)(G 或 T)AC-3′;SduI 识别序列 5′-G(A 或 G 或 T)GC(A 或 C 或 T)C-3′[1]。前者又可表示为 5′-GTMKAC-3′;后者又可表示为 5′-GDGCHC-3′。根据国际生物化学联盟术语委员会 1984 年的命名[1]:非 A 碱基,即 C 或 G 或 T 被命名为 B;非 C 碱基,即 G 或 A 或 T 被命名为 D;非 G 碱基,即 A 或 C 或 T 被命名为 H;非 T 碱基,即 G 或 C 或 A 被命名为 V。这些碱基集合连同以前出现过的 R、Y、M、K、S 和 W 共 10 种。它们在 DNA 群 12 个元素分别作用下的变换关系已经求出[7]。在群元素 R_z 的作用下其变换关系示于图 3.1.6 之中。因为 R_z 对应 T 群绕 z 轴的 180°旋转,因此这个结果亦可通过对图 3.1.3 中绕 z 轴的 180°旋转得到,这里非 A、非 C、非 G 和非 T 碱基分别用△CGT、△AGT、△ACT 和△ACG 表示。在 R_z 作用下,△CGT 和△ACG 互换位置,即 B、V 互换。其他依此类推。特别地,由于 z 轴垂直于图 3.1.3 中的 AT 棱和 GC 棱,故绕 z 轴 180°旋转后,S 与 W 各自保持不变,正如图 3.1.6 所示。

在此项研究的基础之上,作者提出了判断第二类限制性内切酶(简称内切酶)识别多重 DNA 序列的两条规则。这些规则只是判断一多重

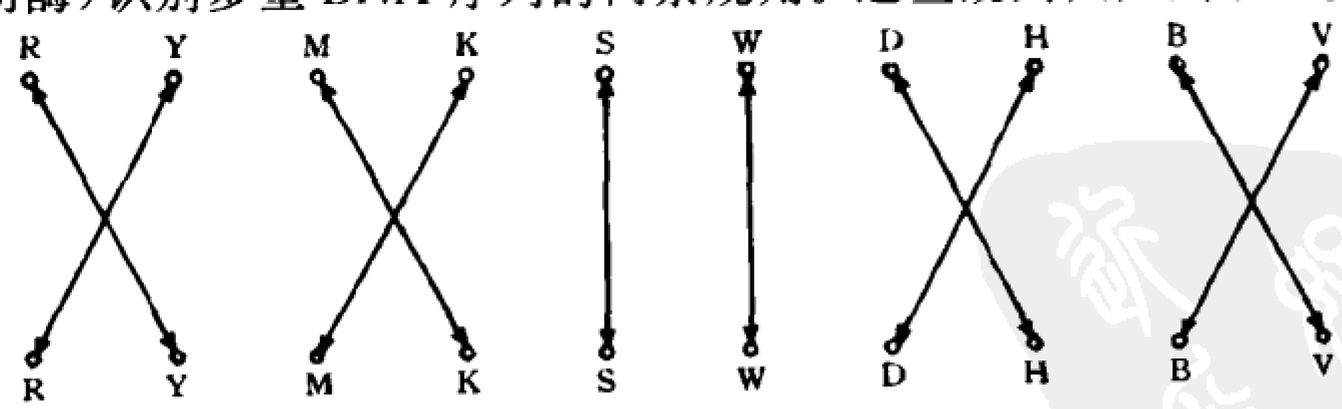


图 3.1.6 10 种碱基集合在互补变换(对应在图 2.1.3 中将正四面体绕 z 轴旋转 180°)作用下的变换关系

DNA 序列是否为一内切酶底物的必要条件，而不是充分的。

定义：对一个由 A、C、G、T、R、Y、M、K、B、V、D、H、S 和 W 共 14 种符号组成的多重 DNA 序列，若对它分别施行 R_z 和 R_f（反向）操作所得到的序列相同，则称此序列为自互补的。

规则 1：所有第二类限制性内切酶识别的多重 DNA 序列必为自互补的。

规则 2：在第二类限制性内切酶识别的多重 DNA 序列中，R 和 Y，M 和 K，D 和 H，B 和 V 必成对出现。S 和 W 可分别成单或成双出现。

我们检查了较新版本的内切酶识别序列表[9]，尚未发现一例是违反以上规则的。根据这些规则，作者进一步预测了一些新的尚未被发现的多重识别序列，如 SGGCCS，GBGCVC，CCSSGG，RGGSCCY 等。但是，自然界中是否存在这样能够识别以上的序列的内切酶，我们不妨拭目以待。

关于 DNA 序列对称性的研究详情，可参见文献[7]。

3.1.3 三种分布的频谱学研究

如果把序列延伸数 n 延拓到正半实轴上，并将其理解为时间轴，则方程(3.1.1)的三个方程成为三个独立的时间过程或时间序列。当然，它们又是 Z 曲线的三个参数方程。于是，关于时间过程、时间序列的一些成熟的数学理论，几乎可以原封不动地引入到 DNA 序列的分析中来。

利用文献[10]中所使用的方法，研究了许多基因序列中的三种分布的频谱分布。图 3.1.7 给出了大肠杆菌 *apto* 基因族（总长度为9 874 bp）中前4 096个碱基的 Z 曲线三个分量的功率谱[11]。在图 3.1.7 中，横坐标为频率 f，单位为1/4 096周。易见，在$f \simeq 1\ 365$处，3 种分布的功率谱都出现了鲜明的峰，尤以 z 分量为甚。该频率的倒数，即对应的周期，等于 $1/1\ 365/(1/4\ 096)=3$，这恐怕是对应三联体密码了。在我们之前，类似的结果，文献上已有多次报道，只不过我们用 Z 曲线的方法将其重新做出而已。但是，我们的方法为解释这一现象提供了新的观察角度。首先，发现非编码区及大量由计算机产生的随机 DNA 序列决不出现这样的峰，而对不同物种的数百个不同基因编码区的计算表明，此峰必定出现[11]。这一结果强烈表明，此峰无疑来自于三联体密码。其次，在 Z 曲线的三个分量中，以 x 或 z 分量的峰最强，y 最弱，甚至不出现。这一结果为解释这一现象提供了新的思路。因为三联体密码并非意味着在编码区的频谱中

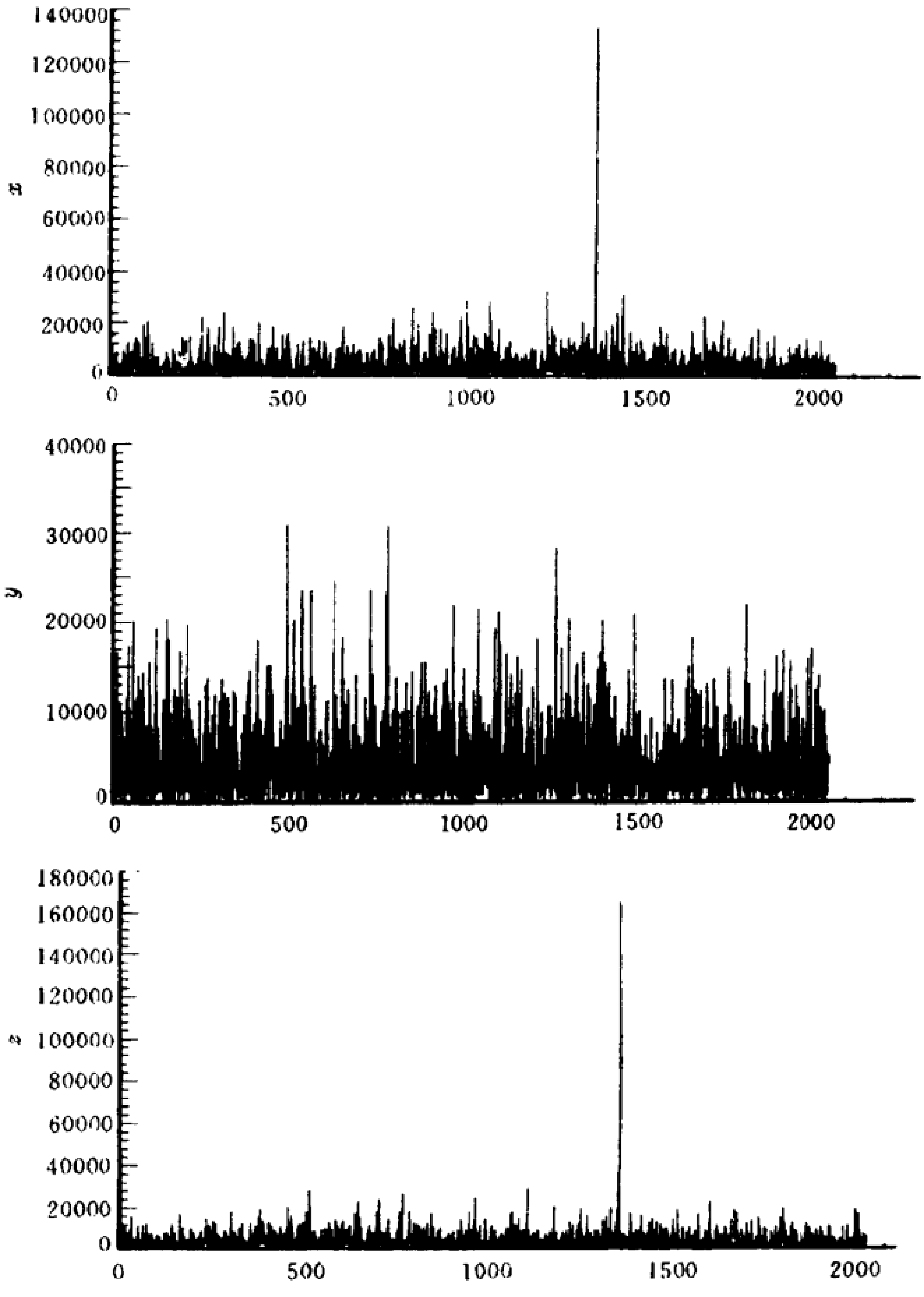


图 3.1.7 大肠杆菌 *apto* 基因族(总长度为 9 847pp 中前 4 096 个碱基所对应的 Z 曲线的 3 个分量经傅里叶变换以后所得到的功率谱

一定要出现周期为 3 的峰,其中必然另有原因,只是目前尚无统一的公认的解释[12]。我们提出以下的解释:对十余个物种近万个基因编码区中碱基频率的分布,用图形法[2~4]进行了分析,结果发现,密码子第 1 位以嘌呤为主,不因物种或基因而变,分析这是由于蛋白质结构的稳定性所要求

的[13]。我们认为这一现象导致了在 x 分量(即嘌呤/嘧啶)的频谱中出现了周期为3的峰。同时发现,在密码子的第三位置上,偏爱的碱基是S/W型的。即有的物种主要偏爱G或C;有的主要偏爱A或T。我们认为这一现象与 z 分量出现的峰有关。这一研究结果正在整理中。

在图3.1.7的三种分布的频谱分布中,除主峰以外,还散布着几乎均匀的各种频率的谱。这一现象在我们所研究过的编码区里普遍存在。这意味着编码序列近似为随机序列。我们这里用完全不同的方法做出了与以前得到的完全相同的结论[14~16]。

需要指出的是,不久前,Voss[17]和 Li - Kakeko[18]也做了DNA序列的频谱分析工作。但前者的分析忽略了DNA序列基本的对称性;后者的工作所分析的量,其生物学意义不明确。而我们的工作不仅充分考虑了DNA序列的对称性,而且三种分布的生物学意义十分明确。

自Peng等人[19]的工作发表以后,DNA序列中是否存在长程相关的问题,引起了广泛的兴趣和讨论。Peng等将DNA序列表示为一维行走的净位移:从原点出发,从第一个碱基算起,若是嘌呤碱基就向负方向走一步;若是嘧啶则向正方向走一步。走了 n 步以后的净位移记为 f_n,这里 $n=1,2,\cdots,L$,其中 L 为序列的长度。在长度为 l 的窗口里计算位移的均方差,然后对全序列求平均,得到所谓Peng函数 $F(l)$。Peng等发现,在 l 的几个数量级范围内,$F(l)\propto l^{\alpha}$,其中对不含内含子的序列包括一些真核生物的cDNA,$\alpha\simeq0.5$;对于富含内含子的序列,$\alpha\simeq0.6\sim0.7$[19]。于是认为后者中碱基存在着长程关联。容易证明,Peng等人的一维行走的净位移 f_n 不是别的,正是 Z 曲线在负 x 轴上的投影,即 $f_n=-x_n$。显然,在Peng等人的工作中,只考虑了三种分布中的一种。而根据我们的理论,任一DNA序列需有三个独立的分布来描述,而仅根据其中一种就对序列的关联性做出判断是不充分的。一个很自然的想法就是把Peng等人的分析方法推广到另外两种分布,即 y_n 和 z_n 中去。最近,我们做了这项工作。对数百个原核和真核生物的基因进行了计算,发现情况远比Peng等报道的要复杂得多,许多基因的三种分布表现了不同的关联行为。结果详见文献[11]。

仔细观察图3.1.5和图3.1.7会发现,三种分布的频谱中都包含了从低频到高频的各种成分,其生物学意义是不同的。高频成分表示了碱基

组成的快速涨落；低频成分表示了碱基组成的大体走势。碱基突变中的有益部分通过自然选择而得以保留，有害部分则被淘汰。而大量中性漂变，由于自然选择对其不起作用也可能被保留下来。我们认为大量保留下来的近期突变对高频成分做出了贡献，而低频成分则是大量突变长期积累的结果，从而保持了相对的稳定性。为了从三种分布的频谱学角度研究生物的分子进化，需要将各种不同的频率成分分离开来。所幸的是，数学家和无线电理论学家已经发展了成熟的数学工具，可为我们所用。我们已将非递归数字滤波理论应用到DNA序列的分析中来。例如，为了得到各种频率成分，可以设计低通、高通、带通和带阻滤波器，其频率响应可根据需要任意设定。关于这类研究在分子进化中的应用，正在大力进行之中。

一个特别值得提出的问题是，以上所讨论的所有频率皆指空间频率，它描述碱基在空间排布的起伏特征。而生物的发育、生长等受到严格的时序调控。那么，生物的时钟频率记录在何处？一般认为是固化在DNA序列之中。我们猜想，这两种频率，即时间频率与空间频率似应有某种对应关系。若能揭示出两者之间的可能联系，将成为发育生物学研究中的一项重要进展。

3.1.4 三种分布的微分几何学研究

方程(3.1.1)中的三个方程是 Z 曲线的参数方程。但实际上，由于自变量 n 是整数，因此 x_n，y_n 和 z_n 可看成是经均匀采样后的离散空间(亦可想象为时间)序列。为了得到连续而光滑的曲线，可采用插值、磨光或最小二乘法拟合等函数论方法来实现。但是，由于所研究的DNA序列都比较长，从算法的通用性和节省计算时间角度考虑，采用基样条函数磨光方法[5]较好。根据样条函数理论，用四阶基样条函数磨光的 Z 曲线，处处存在着连续的一、二、三阶导数。在此基础上，我们可引入微分几何学理论方法来研究 Z 曲线。

描述三维空间曲线最基本的三个几何量是弧长 s、曲率 k 和挠率 τ，它们是建立在曲线的微分基础之上的。要计算弧长，需要曲线的一阶导数存在；要计算曲率，还需要二阶导数存在；而要计算挠率，还要求三阶导数存在。现代研究三维空间曲线几何性质的最重要工具是Frenet标架和Frenet公式。沿着曲线上每一点计算出单位切向量、主法向量和从法向量，如图3.1.8所示。它们是两两正交的单位向量组，可看成是附在曲线

上每一点的右手正交标架。当点移动时,这组标架也随之沿曲线运动,因此它成为曲线上的活动标架。Frenet 标架的一个重要优点是,标架对弧长的微分是由曲率和挠率这两个基本的几何量所描述的(Frenet 公式)。根据三维空间曲线微分几何理论的基本定理,给定了曲率和挠率作为弧长的两个连续可微函数后,曲线可唯一被确定。

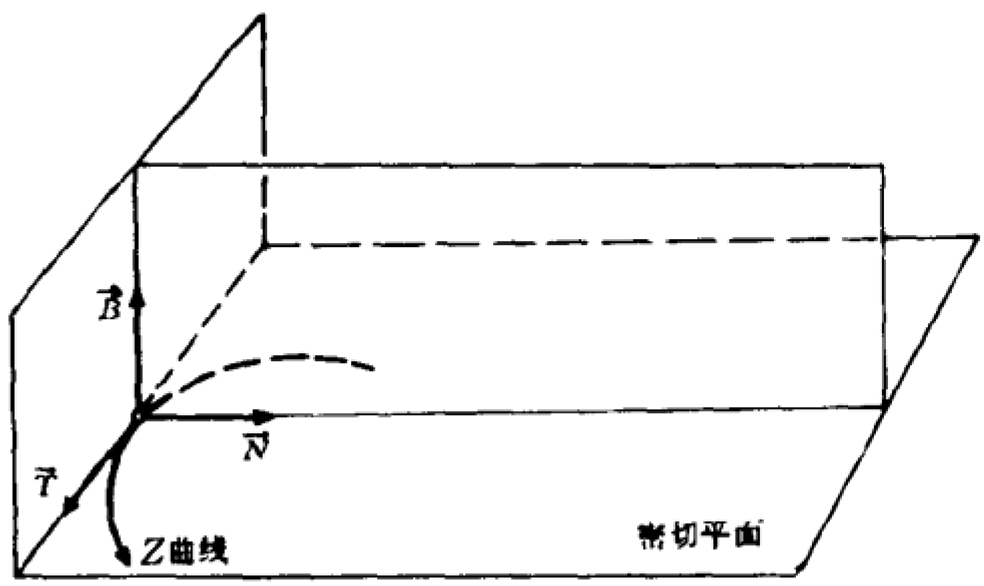


图 3.1.8 Z 曲线的 Frenet 标架

$\vec{T}$、$\vec{N}$和$\vec{B}$分别表示切向量、主法向量和从法向量,三者组成Frenet标架。注意,密切平面最为贴近切点邻域的Z曲线。几何量挠率τ衡量了密切平面法向量的变化率。

以上是微分几何的基本原理,但引入到 DNA 序列中来,可以做一些简化。因为此时弧长可通过简单的办法算出,而不必求积分。Z 曲线是由一系列点两两相连而成(见 § 3.1.1),而任意相邻两点之间的欧氏距离是相等的[5]。因此,弧长 s 与序列延伸数 n 基本上是成正比的。在特殊情况下,即相邻两点用直线相连时,弧长 s 与 n 严格成正比。不失一般性地,不妨假设比例常数为 1。因此,我们利用微分几何理论研究 Z 曲线的基本方程是

$$\begin{cases} s \simeq n \\ k = k(n) \\ \tau = \tau(n) \end{cases} \Longleftrightarrow \begin{bmatrix} x_n \\ y_n \\ z_n \end{bmatrix}, \tag{3.1.5}$$

这一方程表明,曲率和挠率作为 n 的两个函数可以确定 DNA 序列中的三种分布;反之,从三种分布出发,可以求出曲率和挠率两个作为 n 的函数。因此,对于 DNA 序列的研究可以转化为对 Z 曲线的曲率和挠率沿序列的变化的研究。这是利用微分几何学方法研究 DNA 序列的基本思路。

从 Z 曲线出发,求其曲率和挠率的计算比较复杂,拟不在此讨论。问

题是，既然曲率和挠率代表了所研究的DNA序列，那么它们有什么生物学意义呢？我们曾计算了λ噬菌体全部染色体组DNA序列的Z曲线，总长度为48 502 bp。结果发现，在转录从一股跳迁至另一股处，曲率达到一个极大值，同时在此点邻域，序列的碱基组成发生了剧烈的变化。因此可以说，曲率的大小表征了序列中碱基组成变化的剧烈程度。在变化剧烈的区域曲率增大，反之，在组成均匀、变化平坦的区域曲率减小或趋向于零。至于挠率的生物学意义，是与Z曲线的性质及挠率的几何意义密切相关的。众所周知，曲线的密切平面是与曲线关系最紧密相切的一个平面(图3.1.8)。换言之，在所有的切平面中，密切平面同曲线最贴近。对于任意一条平面曲线，在曲线中每一点的密切平面都等同于曲线所在的平面。三维空间曲线的挠率的数值衡量了曲线密切平面的法向量的变化率。粗略地说，也是衡量曲线偏离一个平面的程度。对于一条平面曲线，其挠率处处为零。而根据Z曲线的性质，当一段DNA序列只有两类碱基组成时，它的Z曲线是一条平面曲线[5]。对于原核生物基因的启动子，在转录起始位上游约10 bp附近的pribnow box(TATAAT)，它对应的Z曲线被约束在$x=y, z>0$平面[5]。在此区域Z曲线之挠率处处为零。进入编码区后，由于编码蛋白质稳定性的需要，G、C含量迅速增大[13]。理论上，进入纯G、C区后，相应Z曲线的密切平面将与$x=-y, z<0$平面重合[5]。实际上，在许多基因的编码区中G、C占重要成分。因此可以想象，从启动区到编码区，Z曲线的挠率将发生很大的变化，由近乎零值突变到一个很大的值，然后再减少。初步的计算表明，情况确实如此。总之，Z曲线的曲率和挠率反映了DNA序列中碱基排列的局部模式和总体特征。我们设想，Z曲线曲率和挠率的计算和分析，对于识别DNA序列的不同的功能区等研究，将成为一个常规的有用的工具。

不久前，著名分子生物学家Gilbert在*Nature*上著文指出，当前分子生物学已进入实验与理论并行发展的阶段[20]。他指出，通常认为分子生物学研究的“标准”模式是，问题的解决被认定是纯实验的。“正确”的途径是通过实验识别出某个基因，克隆它，测其序列，研究其产物并进一步通过实验去了解其功能。但是现在情况发生了变化，一种新的研究模式正在出现。当许多甚至某物种的全部基因序列都已测出，并通过存储媒介和计

算机网络广泛传播从而使研究人员容易得到这些序列之后，理论研究的成分将逐步增大，将通过理论方法的洞察力来产生新的生物学知识[20]。对此形势我们应有清醒的认识。

该项研究的目的就是通过数学理论的洞察力来获取新的生物学知识。研究工作全部在国内进行，但以此项研究成果积极参与国际竞争。它系统地开拓了DNA序列分析的几何学研究领域，初步将计算几何学、微分几何学、群的表示理论、时间序列和数字滤波等数学理论引入到DNA序列的分析中来。正如在文章里所看到的，我们已经有了一个可喜的开端，取得了一些有价值的成果，增加了新的生物学知识。但总的来说，整个工作仍处在开始阶段，还有大量工作要做。要特别指出的是，DNA序列的几何学研究是建立在两大基础之上的。它的理论基础是上述的数学理论；它的技术基础是计算机图形学技术。离开了计算机图形学技术的应用，这项研究将寸步难行。这是一项跨越数学、物理学、生物化学和计算机科学的多学科交叉的边缘学科。我们确信，在不久的将来，这项研究有可能发展成一门新兴的分支学科——几何DNA序列学。

参考文献

[1] Cornish-Bowden A. *Nucl Acids Res*, 1985, **13**: 3021

[2] Zhang C T, Zhang R. *Nucl Acids Res*, 1991, **19**: 6313.

[3] Zhang C T, Zhang R. *Int J Biol Macromol*, 1991, **13**: 45

[4] Zhang C T. In: Pickover C A, ed. *Visualizing Biological Information*, Singapore: World Scientific Pub Co, 1993. 83

[5] Zhang R, Zhang C T. *J Biomol Struc Dyn*, 1994, **11**: 767

[6] Yan M, Lin Z S, Zhang C T. *Bioinformatics*, 1998, **14**:685

[7] Zhang C T. *J Theor Biol*, 1997, **187**: 297

[8] Zhang C T. Zhan Y. *J Theor Biol*, 1994, **167**: 161

[9] Perbal B. A proctical guide to molecular cloning, 2nd ed. New York: John Wiley & Sons, 1990

[10] Zhang C T, Cen Y Q. *J Theor Biol*, 1989, **138**: 457

[11] Zhang C T, Lin Z S, Yan M, Zhang R. *J Theor Biol*, 1998, **192**: 467

[12] Shepherd J C W. *Proc Natn Acad Sci USA*, 1981, **78**: 1596

[13] Zhang C T, Chou K C. *J Mol Biol*, 1994, **238**: 1

[14] Zhang C T. *Int J Biol Macromol*, 1989, **11**:9

[15] Zhang C T. *J Theor Biol*, 1990, **142**: 281
[16] Zhang C T, Shang Z X. *J Theor Biol*, 1991, **149**: 257
[17] Voss R F. *Phys Rev Lett*, 1992, **68**: 3805
[18] Li W, Kaneko K. *Europhys Lett*, 1992, **17**: 655
[19] Peng C K, *et al*. *Nature*, 1992, **356**: 168
[20] Gilbert W. *Nature*, 1991, **349**: 99

（张春霆）

§3.2 密码—序列—构象—动力学：一条可能的理论生物学路线

近代生命科学的发展已不单是生物学家的事，需要化学家、物理学家、数学家的通力合作和各门学科的相互渗透。物理学以研究自然界基本规律作为自己的任务，当生命科学已从描述阶段发展到穷究机理的阶段时，自然离不开物理学。

1982—1983年，当笔者的兴趣从传统的理论物理转向生命科学以后，立即发现一个问题：生物学从来就是一门实验科学，有没有理论生物学？如果有，什么是理论生物学？它的地位、作用和价值又如何？1988年笔者在《大自然探索》中试探性地提出“密码—序列—构象—动力学”的反向生物学路线。文中指出：“传统生物学的研究是由形态到细胞，再到分子水平的结构，由宏观生理到微观调控，由表型到基因型。伴随着理论生物物理学的建立，将有可能建立反向生物学。它是由基因入手，由分子序列到分子构象，由结构到功能。从统一原理出发，广泛利用数学工具，定量地阐明千变万化的生命现象。”[1]不久，笔者在 *Nature* 杂志上发现了诺贝尔奖获得者 W. Gilbert 的一段相似的话：“传统生物学解决问题的方式还是完全实验的。而正在建立的新模式是：基于全部基因都将知晓，并以电子技术可操作的方式驻留在数据库中，生物学研究模式的出发点应是理论的。一个科学家将从理论推测出假定，然后再返回到实验中去，追踪或验证这些理论假定。”[2]W. Gilbert 是诺贝尔奖获得者，他的巨大声望给笔者以支持。在这条反向的生物学研究路线中，密码、序列、构象、动力学是四个基本环节。下面是我们近年在这四个环节上获得的若干结果。

1. 密码的进化稳定性

证明现密码表的简并多重态排列是突变危险性极小的；总体排列是突变危险性近似极小的[3,4]。

2. 密码的阴阳对偶性

由表示

U ═══；C ═ ═；G ══ ═；A ═ ═

出发，导出密码表的亲疏水氨基酸分布[5]。

3. 氨基酸突变的统计理论

由核苷酸的随机突变和突变氨基酸在选择压力下的适合度导出氨基酸的突变频率，和实验结果基本相符[6]。

一般说来，反向路线中的密码是指核酸和蛋白质间信息传输中的相互识别关系，它既包括信息从核苷酸序列向氨基酸序列传递的遗传密码，也包括各种调节蛋白质对碱基的识别。这方面还有很多问题值得研究。

4. 核苷酸序列的碱基组成和统计关联与随机序列很接近，生命信息埋藏于强噪声背景中

由编码序列的统计分析得到[7]：

$$\frac{D_1}{H_{\max}} = \frac{H_{\max} - H}{H_{\max}} \sim 0.02, \tag{3.2.1}$$

$$\frac{D_2}{H} = \frac{H - H_{\text{markov}}}{H} \sim 0.04, \tag{3.2.2}$$

$D_1<\epsilon$ 表明，$\sum P_i^2<(1+\epsilon\ln 2)/4<\frac{1}{3}$。$\sum P_i^2<\frac{1}{3}$ 可由直接统计得到[8]。以上(3.2.1)和(3.2.2)两式表明核苷酸序列的噪声特性，这是由于强大的中性突变对进化的影响。如将编码序列分成三个子序列，子序列的 D_1、D_2 值要稍大些，但一般仍接近于原序列的值。偏离大的是第三子序列的 D_3，它比 D_1 大 2～3 倍至 5～6 倍，这是由于同义密码子的非随机作用[9]。

5. 信息参数与进化的相关性

发现编码区存在若干信息参数，它们的类平均与进化呈现相关性[7]。日益增多的核酸资料表明，这种进化参数在粗粒化(类平均)的意义下确实存在，其中描写紧邻碱基关联的 D_2，特别是子序列 $D_2(31)$（密码子 3 位与反密码子 1 位关联），与进化相关性最好[9]。

6. 编码区中的碱基关联的短程为主性，关联最强位点在 0,1 或 2

以上结论是将描述紧邻关联的 D_2 推广到非紧邻关联，定义为 D_k（即

后来一些工作中用的互信息），并引入有限序列的涨落限，对关联长度进行确切定义后得到的[10,11]。这些结果比国外的工作[12,13]早，并更为准确。从生物学角度看，由于核酸序列作为一个整体发挥功能，长程关联是容易理解的，为什么多数编码区具有短程关联？我们对此提出了一种“突变-选择”的理论模型进行解释[14,15]。

7. DNA 行走

我们讨论了 DNA 二维行走，得到了多数编码序列的分维 D_f 为 1.3～1.7，并与进化有粗略相关性[16]。这里 D_f 相当于后来文献[11]中引入的指数 α 的倒数。文献[11,12]考虑一维行走，得到了编码区 $\alpha=\frac{1}{2}$ 因而无关联的结论。我们认为，一维行走不足以判定关联。实际上，从我们的较早工作（二维行走）已经得到了 $\alpha>\frac{1}{2}$，证实了关联的存在[16]。

8. 非编码序列具有不同于编码区的信息参数

对内含子、5′头、3′尾等进行分析容易导出此结论，并看出它们与进化似无明显的相关性[7]。多数非编码区的 D_1 大于编码区，这是由于其中富含多种调控信号所致[7]。最近工作[17]得到了相同结果。

9. 由核酸序列的模糊聚类导出进化树

通过对$\{D_k\}$（$k\leqslant 6$）或其对应概率函数的模糊聚类，可以正确地重建进化树。引入长程关联反使聚类与实验不符[18,19]。由此可见，发展短程关联是对一切物种普遍适合的进化机制。以 rRNA 作为分子钟重建细菌的进化树，还支持了生命进化的三超界学说[19]。

10. 引入最大信息原理，概括编码区的进化机制

由编码序列的统计分析可知，核酸进化受两个因素所支配：碱基随机突变使熵增加；自然选择加强了碱基的统计关联，使马科夫熵减少。因此可用最大信息原理来概括进化机制。令描写碱基关联的某些约束下的信息熵取极大值，可导出每一编码序列的碱基概率分布，大多数序列符合率达 90%以上[20]。

11. 核苷酸序列和基因表达的关系

通过同义密码子的非随机使用以及碱基关联与基因表达水平关系的分析，可以找到编码区的基因表达增强网络位点，我们对大肠杆菌和酵母进行了具体研究。这将有助于改造基因，提高表达[21,22]。

12. 寻找保守位点，研究它们与双螺旋结构局部偏差的关系

序列与进化，序列与表达，序列与结构和功能，是研究核酸序列的三个目的。研究保守位点与双螺旋结构局部偏差的关系，有助于了解蛋白质对核酸功能位点的识别[23]。

13. 蛋白质二级结构预测的统计力学基础

由 Zimm - Bragg 模型讨论二级结构经验预测的统计力学基础，证明了二级结构和二肽频率的相关性强于单肽频率 4 倍，因此好的经验预测必须以二肽频率资料为基础[24~27]。

14. 蛋白质折叠(三级结构)的动力学途径依赖性

新生肽链边生长边折叠的格子模拟表明，它能较快地达到折叠态，比完整肽链从无规线团到折叠态快得多。这提供了折叠的动力学途径依赖性的一个实例，也表明折叠的高速度必须是动力学的[28]。

15. 构象局域态与构象势不对称性的关系

从分子水平研究生命过程的动力学，基本变量是构象坐标（主要为扭角）和前沿电子。为了建立构象-电子系统的量子理论，首先必须定义构象局域态。我们证明了当势不对称性大于 $10^{-2}\sim10^{-1}$，全部构象态都是局域的。基态的局域性出现更早，只需 $10^{-6}\sim10^{-3}$的不对称性[29,30]。

16. 建立大分子的构象动力学

用非绝热算符方法建立构象-电子系统量子跃迁的理论形式，可用来讨论无辐射跃迁、施主受主电子跃迁、光致构象电子跃迁等[29,31]。

17. 用构象动力学讨论蛋白质折叠速率

导出折叠时标 10^{-3}秒，速率比例于 $I^{-\frac{5}{2}}$和 $n^{\frac{3}{2}}$等，I 为参与扭转的原子团惯量矩，n 为参与一次跃迁的扭角数。据此可解释 α 螺旋、β 折叠等规则构型在折叠中较早出现，还可据此提出分子伴侣的理论模型[32]。

18. 链式分子的构象——电子运动

由构象动力学导出包含新传播项的推广的 Davydov 方程[33]，研究构象振动对合作效应的影响[29,31]。

19. 复制动力学

提出 DNA 解旋是拓扑量子跃迁，复制叉运动速度受酶扩散速率所限制[34]。

20. 提出大分子溶液的热滞理论

解释鱼抗冻多肽的冰晶生长抑制现象，这提供了分子组装发挥功能的一例[35,36]。

以上工作有的还只是初步结果，主要是想说明我们在"密码—序列—构象—动力学"这条路线若干环节上的探索，相信只有经过此路线上进一步的努力，才能真正地揭示生命的本质。

参考文献

[1] 罗辽复. *Exploration of Nature*(大自然探索), 1988, **7**(4): 9;(物理) *Physics*, 1991, **20**(2): 85

[2] Gilbert W. *Nature*, 1991, **349**: 99

[3] Luo L F. *Origins of Life*, 1988, **18**: 65

[4] Luo L F. *Origins of Life*, 1989, **19**: 621

[5] 罗辽复. *ASNUI*(内蒙古大学学报), 1992, **23**: 395

[6] Luo L F, Zhou Y M. *Origins of Life*, 1990, **20**: 419

[7] Luo L F, Tsai L, Zhou Y M. *J Theor Biol*, 1988, **130**: 351

[8] Zhang C T, Zhang R. *Int J Biol Macromol*, 1991, **13**: 45

[9] 罗辽复,胜利. *ASNUI*,1990,**21**: 229

[10] Luo L F, Li H. *Bull Math Biol*, 1991, **53**: 354

[11] Luo L F, Tsai L, Lee W J. *Phys Rev*, 1990, A**41**: 5441

[12] Peng C K, *et al*. *Nature*, 1992, **356**: 186

[13] Buldyrev S V, *et al*. *Phys Rev*, 1995, E **51**: 5089

[14] Herzel H, Grasse. *Physica*, 1995, A**216**: 518

[15] Luo L F, Trainor L E H. *J Theor Biol*, 1992, **157**: 83

[16] Luo L F, Tsai L. *Chinese Phys Lett*, 1988, **5**: 421

[17] Mantegna R N, *et al*. *Phys Rev Lett*, 1994, **73**: 3169

[18] 纪丰民,罗辽复. *ASNUI*,1992, **23**: 194

[19] Luo L F, Ji F M, Li H. *Bull Math Biol*, 1995, **57**: 527

[20] Luo L F, Bai G Y. *J Theor Biol*, 1995, **174**: 131

[21] 李宏,罗辽复. *Acta Biophysica Sinica*(生物物理学报), 1995, **11**: 237;*J Theor Biol*, 1996,**181**:111

[22] 罗辽复,李宏. *ASNUI*,1995, **26**(6):690

[23] 胜利,罗辽复. *ASNUI*,1991, **22**: 234

[24] 罗辽复,董玉枝. *Chinese Biochem J*(生物化学杂志), 1988, **4**: 173

[25] 李晓琴,罗辽复. *ASNUI*,1992, **23**: 534

[26] 罗辽复. *ASNUI*,1993, **24**: 21

[27] 罗辽复. *Chinese Biochem J* (*special issue*),1993: 6

[28] 李炜疆,罗辽复. 中国学术期刊文摘, 1995 年增刊;罗辽复.*ASNUI*, 1995, **26**: 172;

1996,**27**:333

[29] Luo L F. *Int J Quant Chem*, 1987, **32**: 435

[30] 董玉枝,周宇明,罗辽复. *ASNUI*,1986, **17**: 711

[31] 罗辽复,栾祺浩,马克健,王建英. *Acta Biophys Sinica*, 1985, **1**: 172

[32] Luo L F. *Int J Quant Chem*, 1995, **54**: 243

[33] 罗辽复. *ASNUI*, 1990, **21**: 506

[34] 罗辽复. *ASNUI*, 1993, **24**: 282; 1995, **26**: 496

[35] Li Q Z, Luo L F. *Chem Phys Lett*, 1993, **216**: 453

[36] Li Q Z, Luo L F. *Chem Phys Lett*, 1994, **223**: 181

(罗辽复)

§3.3 RNA结构的语言与代数拓扑特征

生命遗传是通过DNA信息的世代传递而进行的。笔者在“RNA在分子生物学研究中的重要地位”一文中已指出:中心法则所表达的信息传递,既反映DNA(有时也可能是RNA)作为原初遗传信息载体的核心地位,又说明了RNA作为信息传递中间环节的重要性,以及信息传递的方式和途径。人们受到生物遗传信息传递的精确性和高效率的启示,总是希望能够找到某种类似人类信息交流的“语言”一样的东西来揭示信息传递及其表达调控的深刻内涵。这种看法也很自然朴素,因为形式语言已经有了它的规律与理论[1]。试图从形式语言的角度来探讨遗传信息的相互作用方式可能是一条有意义的途径。

语言理论的数理研究始自乔姆斯基(Chomsky),他与沃森和克里克著名的双螺旋的发现是同时代的。二者有可能相互融合、交叉,彼此推动发展。

在1994年的香山会议上我们已经提出这种初步看法。近一年来,云南大学非线性复杂系统研究中心,现代生物中心与中国科学院昆明动物研究所合作对DNA信息传递用语言理论进行研究。这里对这个主题作概述性汇报。

3.3.1 RNA的二级结构语言的简要回顾[2]

1994年,我们首次提出用Dyck语言描述RNA的结构问题,其内容如下:

1. 文法的化学键本质

RNA 以单链形式存在，通过回折为局部螺旋，再与突环相间排列形成复杂的特征性二级结构。在 RNA 的二级结构中，包含茎区、环区和夹泡等要素[3]。从拓扑的观点来看，这些二级结构只要用两种要素，即茎区与环区（半环或开环），就可构造出来。茎区有很强的吸引力，可形成规则的结构。环区与茎区正好相对立，它们在整体上有强的排斥作用，二者都是对自由能有重要贡献的位形。

茎区的结构很简单，从化学键本质看，沃森-克里克氢键所形成的配对可以称为氢键文法

$$\mathrm{H} \to a\mathrm{H}a^* \mid g\mathrm{H}g^* \mid c\mathrm{H}c^* \mid t\mathrm{H}t^* \mid,$$

这组产生式将直接生成 RNA 结构中的茎区(stem)。茎区产生式是由 Searls 给出的[4]，其化学含义的解释则是由我们给出的。

环区的键结构很复杂，从数学上我们可以用氢键文法的补集来构成，称它为游离键文法。

$$\overline{\mathrm{H}} \to a\overline{\mathrm{H}}\neg a^* \mid g\overline{\mathrm{H}}\neg g^* \mid c\overline{\mathrm{H}}\neg c^* \mid t\overline{\mathrm{H}}\neg t^* \mid,$$

$$\neg a^* = \{V_T \backslash a^*\} = \{g, c, a\},$$

$$\neg c^* = \{V_T \backslash c^*\} = \{t, c, a\},$$

$$\neg g^* = \{V_T \backslash g^*\} = \{t, g, a\},$$

$$\neg t^* = \{V_T \backslash t^*\} = \{t, c, g\}。$$

游离键展开之后包含排斥键与弱吸引键(如 Hoogsteen 氢键)

$$\overline{\mathrm{H}}_1 \to \text{purine } \overline{\mathrm{H}} \text{ purine} \mid \text{pyrminidine } \overline{\mathrm{H}} \text{ purine} \mid$$
$$\text{pyrminidine } \overline{\mathrm{H}} \text{ pyrminidine}。$$

由这类键文法产生式所构造出来的环区，其物理本质是对-对即单对排斥游离句法。而真正强游离句法，则应具有远程的排斥条件，即二对排斥、三对排斥、……直到 n 对排斥。从终极符号的关联来看，对排斥是上下无关，但是高于对排斥条件的文法，将可以依赖上下文。在环区，整体上存在强排斥作用，而在局部可能存在弱吸引作用。所以从理论上有怎样的结构，并使它的热力学自由能也相当稳定，需要较高深的文法来讨论。关于这一点我们将在今后针对生物学问题再加以细致研究。估计对-对游离句法是大体符合生物实际的。

我们来讨论环区碱基数为奇数的修补句法。

茎区是碱基的成对组合。其碱基总数是偶数，环区的碱基数将是自由

的,可以是偶数个或奇数个。当它是偶数个时我们可采取空元:

$$A \rightarrow \varepsilon,$$

当被描述者是奇数个时,我们应在文法使用中加以单元修补,使之成为奇数。这个文法的化学含义在本质上是磷酸二酯键,即

$$\mathrm{A} \rightarrow a_1。$$

但是单元修补句法存在一些问题:因为 a_1 是自由的,可能与已形成的环配对成茎,所以需要考虑按上下文作出特定修补。

综上所述,我们已经讨论了磷酸二酯键(A),沃森-克里克氢键(H),强排斥键($\overline{\mathrm{H}}$)与弱吸引键 Hoogsteen 氢键($\overline{\mathrm{H}}_1$)。这些键都是化学键,它在键文法中表示非终结变元,我们可以统一概括为 $\mathrm{S}(\mathrm{A}、\mathrm{H}、\overline{\mathrm{H}}、\overline{\mathrm{H}}_1)$。统一概括的结果在文法作用上将使茎区与环区的生成有统一的文法表述,二者可以相互转换而产生一个完整的茎环结构单位。

如果进一步考虑有两个或多个茎环结构单位彼此并列生长,则需要如下的分裂句法:

$$\mathrm{S} \rightarrow \mathrm{SS},$$

这个句法是非线性的,它具有复杂的功能。

我们将所讨论的生成茎环结构的全部产生式 P 总结起来,这样获 RNA 二级结构文法为:

$$\mathrm{S} \rightarrow {}^{①}\mathrm{SS}\,|\,{}^{②}a_i\mathrm{S}a_i^*\,|\,{}^{③}a_i\mathrm{S}\neg a_i^*\,|\,{}^{④}a_i\,|\,{}^{⑤}\varepsilon, \qquad (\Omega)$$

这样的语法我们赋给它相应的名称:

① 分裂句法 S_1;

② 茎区句法 S_2(嵌套句法 S_2);

③ 环区句法 S_3(对游离句法 S_3);

④ 环终结句法 S_4;

⑤ 空串 S_0。

全部 P 产生式可以产生多种多样的茎环结构,我们概括地称它为 Ω 语法。这个语法用来描写 RNA 的各种二级结构在原则上是有意义的。

2. Dyck 语言的确定

Ω 文法是属于上下文无关的文法,只要求产生环的游离句法是采用单对排斥或单对弱吸引句法。如果要求高于单对吸引,则文法可能高于上下文无关。对于上述文法(Ω)乔姆斯基范式存在,可表达为下列形式的产

生式：

$$A \to BC, \quad A \to a,$$

从语言理论知道[5]

定理 1：$L(\Omega) \in L_2$。

它是乔姆斯基阶层文法中的第二类型文法——上下文无关文法。非确定论下推自动机所接受或识别这类语言。这是一个有意义的结果，说明 RNA 二级结构及其相应的这类遗传语言确有自动机的本质，它或多或少地以不确定论的形式支配着符号语言的产生与信息的传递与转换过程。

3. Ω 语言代数

作为基因信息转录体的 RNA 序列的碱基数一般在 $10^2 \sim 10^4$ 之间。虽然有了文法，要产生这样长的序列语言，需要应用千百次的产生式来推导生成，这是很烦琐的。若我们将字母上升为词汇，由机器语言归约为汇编语言（与计算机由源程序编译为可执行的目标程序的道路相反，这里实质上研究的是反汇编语言），那么更易把握语言本质。不很严格地说，这样做是归约为生成词的“抽象自动机”。

首先若用 n 次嵌套句法 S_2：

$$S \to \omega S \omega^+,$$

$$b_i \in V_T, \qquad \omega = b_1 \cdots b_n,$$

这里 $\omega^+ \equiv [(b_1 b_2 \cdots b_n)^R]^* = \overline{\omega^R}$。表示词 ω 的厄米共轭：镜象 R 与氢键共轭 * 两种运算的合成。这样将形成茎区的嵌套。

其次若用 $(m-1)$ 次游离句法，再用一次修补句法 S_4（若环区碱基为奇数），或者再用一次空串 S_0（若环区碱基为偶数），

$$S \xrightarrow[S_4]{(m-1)S_2} \beta,$$

$$S \xrightarrow[S_0]{(m-1)S_2} \beta,$$

这里 β 词表达了 RNA 二级结构的环区。

至此，我们找到了 Ω 语言的基本句法，即表达式 Ω，它的基本单元是核酸大分子表达的生物的回文[6]。

$$\Omega = \omega \beta \omega^+。$$

回文是下推自动机语言中重要的角色，这里生物回文只是镜象加氢键共

轭，并不会改变语言本质的阶层。于是就有了汇编语言$(\omega\cdot\beta)$。

在汇编语言$(\omega\cdot\beta)$的基础上，我们提升抽象编译语言。考虑分裂句法与嵌入句法之间使用次序的先后，因为语言的推导一般是不可对易代数，因而序是重要的。

若先使用 S_1 后使用 S_2

$$S \xrightarrow[S_1]{} SS \xrightarrow[S_2]{} \omega_1 S\omega_1^+ \omega_2 S\omega_2^+,$$

若再使用 $S\to\beta$ 并定义为：

$$\Omega\cdot\Omega=\omega_1\beta_1\omega_1^+\omega_2\beta_2\omega_2^+,$$

这时两个核酸回文结构是并联相接的，我们称为点乘。

若先使用 S_2 后使用 S_1，再使用 S_2

$$S\to\omega_1 S\omega_1^+\to\omega_1 SS\omega_1^+\to\omega_1\beta_1\omega_2\beta_2\omega_2^+\omega_1^+\to\omega_1\omega_2\beta_2\omega_2^+\beta_1\omega_1^+,$$

定义两个核酸回文结构相嵌(串)联接为花乘

$$\Omega_1 * \Omega_2=\begin{cases}\omega_1\beta_1\omega_2\beta_2\omega_2^+\omega_1^+\\ \omega_1\omega_2\beta_2\omega_2^+\beta_1\omega_1^+\end{cases},$$

这种花乘可以通过定义左嵌入与右嵌入来解决。同时也可以用定义半 Dyck 集的方法来确定这种乘法的唯一性。不难看出任何核酸的二级结构都是以 Ω 语言为单元生成的终极式，因此核酸的二级结构都能由以 Ω 为元素下的点乘与花乘的代数运算来描述。由于基本 Dyck 语言属于上下无关语言 L^2，则我们有下列定理。

定理 2：词 Ω 在点乘与花乘下是代数封闭的。

这里的代数封闭包含下列含义：

① 并接(即点乘本身)；② 并；③ Kleene 闭包；④ 代换；⑤ 同态。

这里封闭的生物含义是多种的，但表明无论如何嵌与并接其二级结构都不超出 Dyck 范畴。进一步揭示这些代数性质是我们的重要方向之一。

4. Ω 语言代数的拓扑几何图

如果说 Ω 语言代数运算已形成了我们的编译语言，那么它已比较接近我们对生物的“计算表达”，这个语言应揭示所计算对象的几何结构。下面，考察一个二级结构单元 Ω 的几何意义：

$$\Omega=\{\omega\omega^+\cdot\beta\},$$

$$\begin{Bmatrix} \text{环}, & \text{茎}, & \text{联接区} \\ (\text{中区}) & (\text{头区}) & (\text{足区}) \end{Bmatrix}。$$

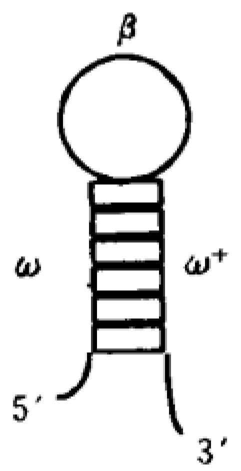
(1) Ω 的并接：两个茎环 Ω_1，Ω_2 的足区与足区并接 $\Omega_1\Omega_2$。

我们沿用 $5'\to3'$顺时针阅读并规定第二个 ω 为具体厄米共轭词，则有

$$\Omega_1\Omega_2=\omega_1\beta_1\omega_1^+\omega_2\beta_2\omega_2^+=\Omega_1\cdot\Omega_2,$$

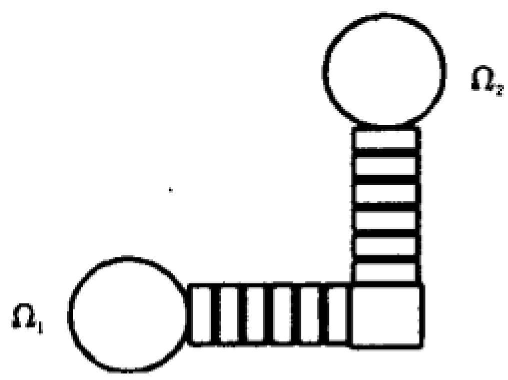
这说明点乘“·”的几何意义为足区与足区并联。

(2) Ω 的嵌接：一个茎环 Ω_2 的足区嵌入 Ω_1 的环区，定义为 Ω 嵌接 $\Omega_1*(\Omega_2)$。我们沿图读出 $\Omega_1*(\Omega_2)=\omega_1\beta_1\omega_2\beta_2\omega_2^+\omega_1^+\equiv\Omega_1*\Omega_2$ 即头足嵌接正是花乘“*”，既然所有 RNA 二级结构的 Ω 代数语言表达了它的全部序列的复杂结构，则任何多个茎环 Ω_1 的复杂连接图可以用点乘与花乘唯一表达，如图 3.3.1、3.3.2、3.3.3 所示。



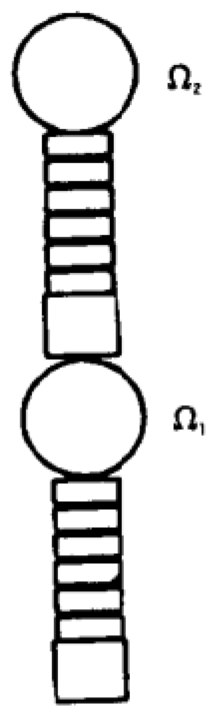
Ω

图 3.3.1 RNA 二级结构单元



$\Omega_1\cdot\Omega_2$

图 3.3.2 点乘两个 Ω 单元的并接



$\Omega_1*\Omega_2$

图 3.3.3 花乘两个 Ω 单元的嵌接

5. 语法派生树

我们已有了 RNA 的语法规则，任何具体的 RNA 二级结构都可以由编译语言(Ω)通过点乘与花乘直观地写出来，这里我们用脊椎动物 5.8 S rRNA的一般模型作为例子来研究基因序列如何从“编译语言”推导到“机器语言”，在推导过程中，Ω 二级结构的语法派生树是一个重要的代数对象。我们需要注意，对于 5.8 S rRNA 从结构上有下列的几何直观

(图 3.3.4)。其 Dyck Ω 模式为 $\Omega_1 * (\Omega_2 * (\Omega_3\Omega_4)\Omega_5)$，上述推导所对应的语法派生树为：

$$
\begin{aligned}
S &\longrightarrow \omega_1 S\omega_1^+ \\
&\longrightarrow \omega_1 SS\omega_1^+ \\
&\longrightarrow \omega_1 SSS\omega_1^+ \\
&\longrightarrow \omega_1\beta_1\omega_2 S\omega_2^+\omega_5\beta_5\omega_5^+\omega_1^+ \\
&\longrightarrow \omega_1\beta_1\omega_2 SS\omega_2^+\omega_5\beta_5\omega_5^+\omega_1^+ \\
&\longrightarrow \omega_1\beta_1\omega_2\beta_2 SS\omega_2^+\omega_5\beta_5\omega_5^+\omega_1^+ \\
&\longrightarrow \omega_1\beta_1\omega_2\beta_2\omega_3\beta_3\omega_3^+ S\omega_2^+\omega_5\beta_5\omega_5^+\omega_1^+ \\
&\longrightarrow \omega_1\beta_1\omega_2\beta_2\omega_3\beta_3\omega_3^+\omega_4\beta_4\omega_4^+\omega_2^+\omega_5\beta_5\omega_5^+\omega_1^+ 。
\end{aligned}
$$

从语法派生树很清楚地看到，从树根 S 生长出一株的全过程，它的叶子集构成在汇编语言水平上的终极符号，它的树干或内结点是非终极符号。在汇编语言(ω,β)的基础上 Ω 语言的树结构不超过 3 叉树，在“机器语言”(a,g,c,t)的水平上，树结构也不超过 4 叉树。在平面树图上某一枝干分叉是严格有序的。若依从左到右，或反时针方向，它们有左右序。按照这个序，* 积的左嵌入 β 或右嵌入 β 分别在左枝或右枝。

每棵语法派生树的叶子集从 5′到 3′依序生成 RNA 序列，反之叶子集是否对应唯一的语法树呢?这是一个重要而复杂的问题，我们将在不同的水平上回答与讨论。

6. Dyck 推导树、树干和树叶

对于一个树结构，它有很多概念，如树叶的水平和树叶所在树干的深度，树的节点通路长度等，我们将借鉴这些理论来研究 RNA 的语法树结构。这里最优先研究的是 Dyck 推导树的树变换理论。

在未讨论树变换之前，我们预先在两种水平上，回答叶子集是否唯一对应一棵语法树。这个问题之所以重要，是因为它的回答是判定 Ω 语言是否有歧义语法的充要条件。

首先，在(ω,β)“汇编”语言的水平上来回答，如果 ω_1 与 β_1 是完全可

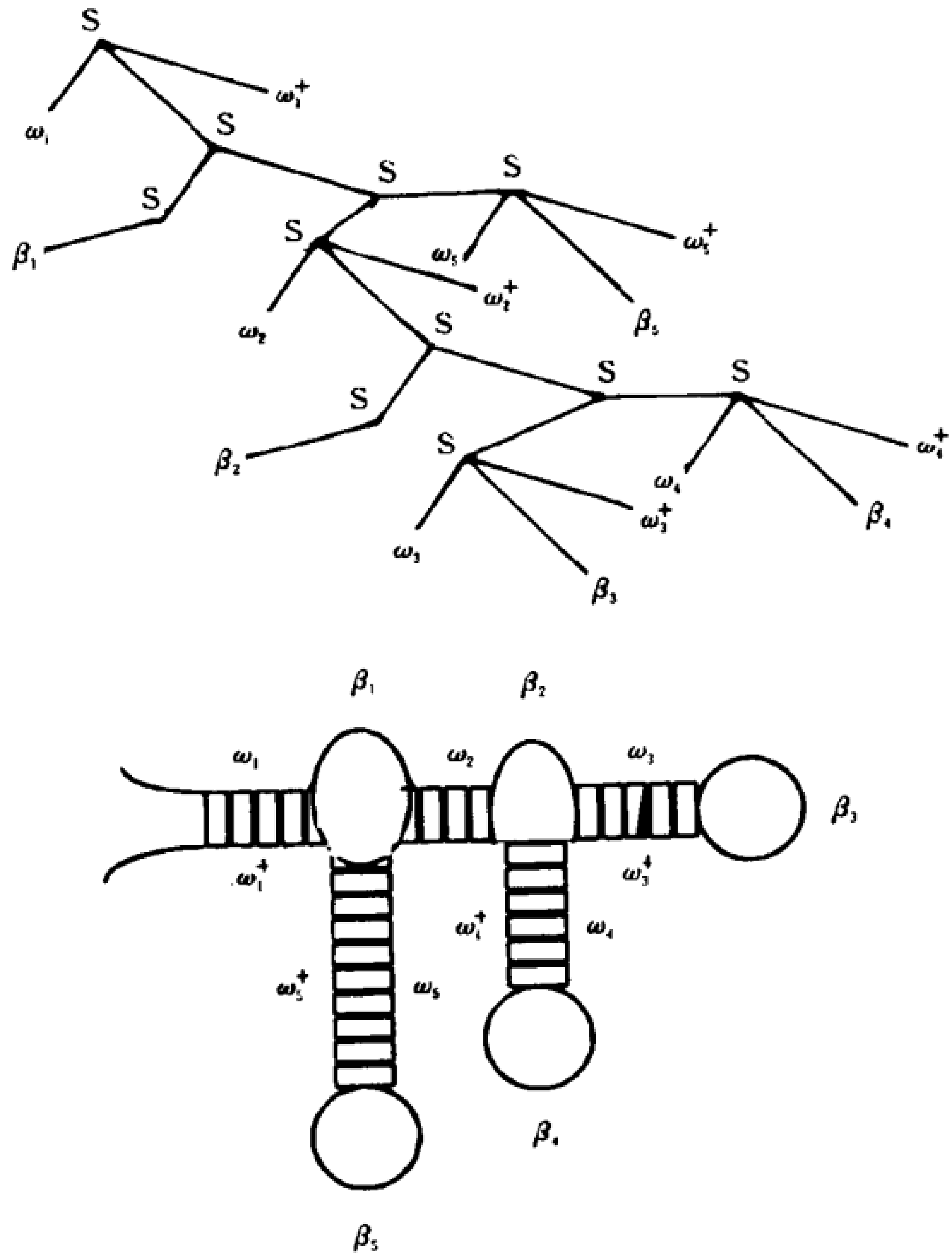


图 3.3.4 5.8 S rRNA 二级结构图与语法派生树

区分的，则它们是唯一的。

其次，若在(a,g,c,t)机器语言的水平上来回答，则它们不是唯一的，所以 Dyck 语言的机器语言是对应不确定论下推自动机，一个$\{a,g,c,t\}$序列将可以对应相当多的语法推导树，要解决这个问题，还需要在$\{\omega,\beta\}$层次上引入语法学的操作。

7. **Dyck 树变换**

一个完整的树图与一个完整的弧序列相对应，这是树的所谓括号表示。以 5.8 S rRNA 为例，它的语法树表示依照树干的水平进行括弧表示：

$$\mathrm{S}((\omega_1)(\mathrm{S})((\omega_2^+))$$

$S(\omega_1 S(S(\beta_1), S(S(\omega_2 S(S(\beta_2), S(\omega_3, S(\beta_3), \omega_3^+), S\omega_4 S(\beta_4), \omega_4^+), \omega_2^+, S(\omega_5, S(\beta_5), \omega_5^+), \omega_1^+)$,

括弧的层面代表推导的深度，我们看见每一个左括弧都由S的推导而来，即非终极符号都由括号代表

① 我们可以省去S，即 S(→(

$(\omega_1(\beta_1)((\omega_2((\beta_2),(\omega_3(\beta_3)\omega_3^+)(\omega_4(\beta_4)\omega_4^+))\omega_2^+))(\omega_5(\beta_5)\omega_5^+))\omega_1^+$；

② 如果我们再加强缺省，将$(\omega_1 \rightarrow (_1, \omega_1^+ \rightarrow)_1$

$(_1((\beta_1)((_2((\beta_2))(_3(\beta_3))_3(_4(\beta_4))_4))_2)(_5(\beta_5))_5))_1$；

③ 再进行简化抽象 $(\beta_1) \rightarrow \beta_1$

$(_1(\beta_1(_2(\beta_2)(_3\beta_3)_3(_4(\beta_4)_4))_2))(_5\beta_5)_5))$；

④ 再略去由于花乘嵌入β的无字括号，它只代表变元或内节点，它与叶子无关，这样

$(_1\beta_1(_2\beta_2(_3\beta_3)_3(_4\beta_4)_4)_2(_5\beta_5)_5)$；

⑤ 这样最后抽去足标，也抽去β，获得完整的叶子集

((( )( ))( )),

这个集合正是点乘和花乘的组合

$$\Omega * (\Omega * (\Omega \cdot \Omega)\Omega),$$

如果我们将括弧只赋给花乘，点乘用字母或词标记，则有

$$((\Omega_3\Omega_4)\Omega_5),$$

这是真正的括弧文本语言。

注意我们省略过程的每一步，如果可以完全地唯一地恢复全部推导过程再现语法派生树，那么我们就很清楚地回答了一个问题，在(ω,β)(汇编)语言水平上，自上而下是无歧义的。但是在(a,g,c,t)水平上将会出现严重歧义，这个问题非常复杂，留待今后研究。

如果要解决唯一性，可将所有的省略步骤，通过树变换转移为子集，由叶子集与实际RNA 序列集对应。那么我们希望所有的省略步骤所代表的变换是一一对应的，于是这个唯一性问题就可能在(ω,β)语言水平解决。

3.3.2 Dyck 文法的功能

我们在前面获得的 Dyck 文法只是一个借用自动机语言描写 RNA 二级结构的形式框架。即使这个数学框架还比较粗糙，甚至拓扑化，但它

已能在远距离上把握住二级结构的基本特征并将导致对 RNA 描述上的重要应用。目前，这些应用中首先可以展开的研究是 Ω 模式的识别与分类。其次是 Ω 模式识别基础上的语法分析器。其他更深入的包括 Ω 模式语义及其算法等相当有兴趣的问题，有待研究深入之后才能有条件讨论。

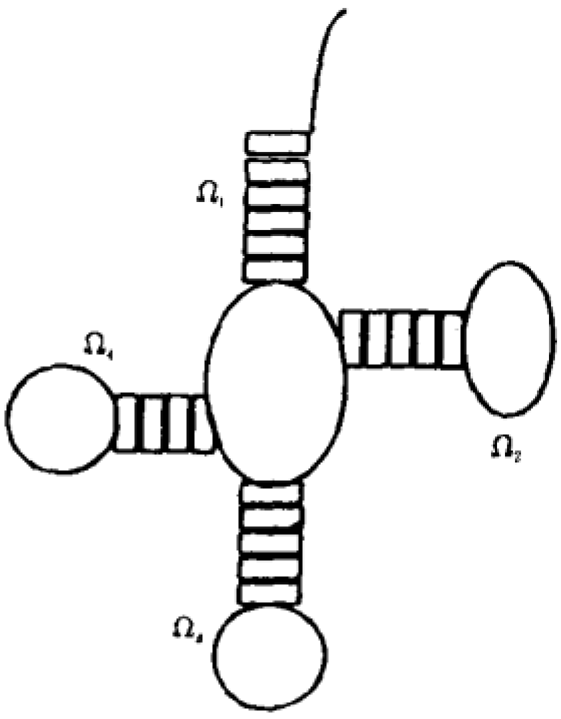


图 3.3.5 三叶草 Ω 模式

1. *Ω 模式识别——阅读规则*

从可嵌入二维平面几何的拓扑观点看，每一 RNA 的核苷酸序列，如果它已形成生物遗传信息传递过程中的某种茎环二级结构（这里暂时不考虑三级以上的复杂结构），那么，它都唯一地对应着一个 Ω 模式，即我们在汇编 Ω 语言的水平上没有歧义。

这样我们可以建立阅读规则：

① 整体沿顺时针方向从 5′到 3′连续阅读，顺序以 i 下标记录；

② 第一次遇到的茎边记 ω_i，以后遇到它的配对的茎边记为 ω_i^+，遇到环状部分记为 β_i 或 β_i', β_i''…等；

③ 如果有局部反时针出现时，仅茎对(ω_i, ω_i^+)共轭换位为(ω_i^+, ω_i)。

在这个阅读规则下每一个 RNA 二级结构都有唯一的 Ω 语法结构，即它的(ω, β)序列，反之，不同的 Ω 序列将对应不同的二级结构，这是二级结构模式识别的基础。

2. *Ω 语言的代数模式的识别*

代数表达本身就是一种句法模式，因此它的识别是自然的，我们给出几个简单的例子来说明。

① 三叶草模型

Ω 代数模式：$\Omega_1 * (\Omega_2 \cdot \Omega_3 \cdot \Omega_4)$，

Dyck 文法模式：((0)(0)(0))。

② 5 S rRNA 的 Fox 模型

Ω 代数 Ⅰ 型　$\Omega_1 * (\Omega_2 \cdot (\Omega_3 * \Omega_4))$；

　　　Ⅱ 型　$\Omega_1 * (\Omega_2 * \Omega_3 \cdot \Omega_4 * \Omega_5)$。

Dyck 文法：

　Ⅰ 型：(( )(( )));

II 型：((( ))(( )))。

③ 大肠杆菌 16 S rRNA 的二级结构(Ω 代数：中心区)：

$\Omega_1 * (\Omega_2 * (\Omega_3 * (\Omega_4 * (\Omega_5 * (\Omega_6 * (\Omega_7 * \Omega_8))))),$

$\Omega_9 * (\Omega_{10} * (\Omega_{11} * (\Omega_{12})))\Omega_{13}) * (\Omega_{14} * (\Omega_{15}))\Omega_{16} * (\Omega_{17}))$。

也有相应的 Dyck 括弧文本表示。

3. 模式的分类

模式识别的功能将所有的二级结构化成两种符号序列：

代数：A=Ω·，　　　B=Ω*；

文法：A=( )( )，　B=(( ))。

从拓扑角度，平面二级结构的意义上，一切 RNA 序列是 A，B 两种序列及其子序列的组合。

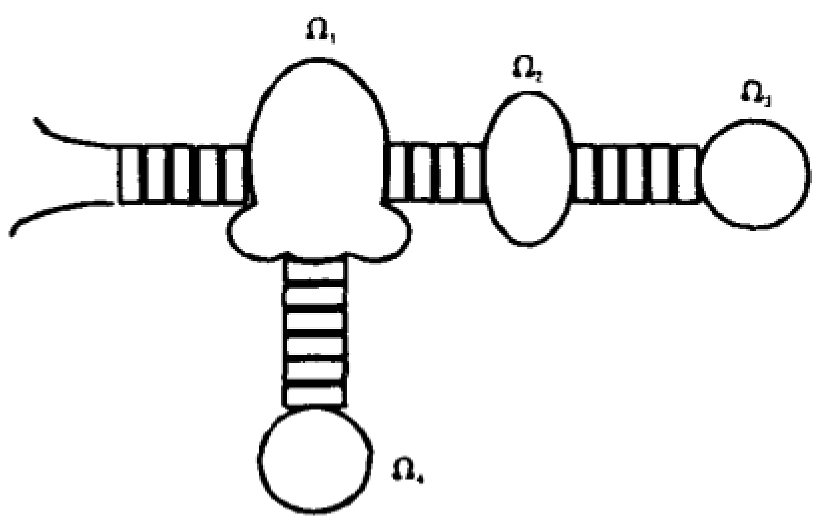


图 3.3.6 原核生物 5 S rRNA Fox 模型 I 的 Ω 模式

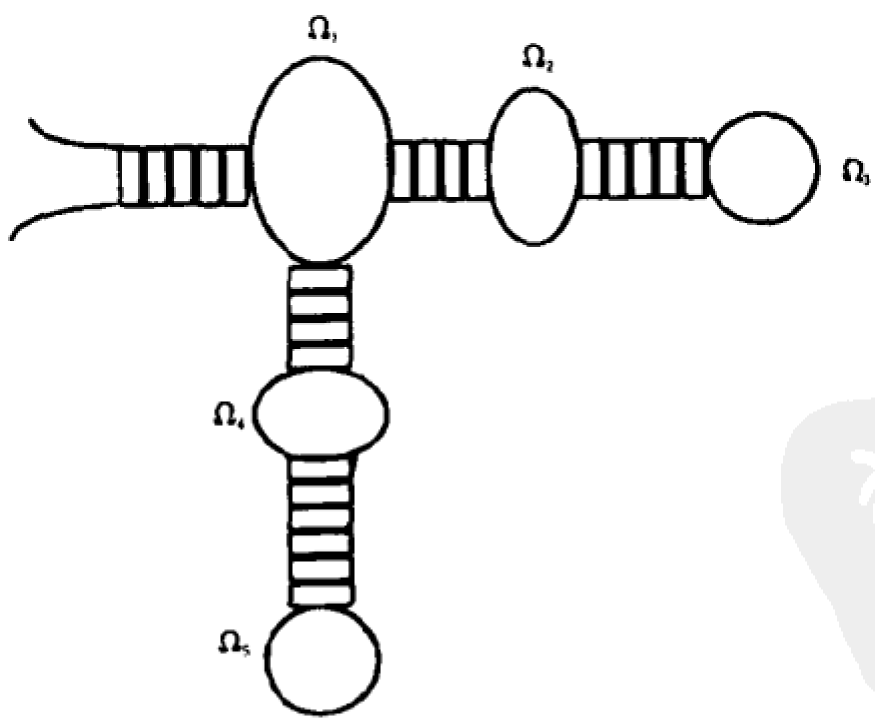


图 3.3.7 真核生物 5 S rRNA Fox 模型 II 的 Ω 模式

相似结构的A,B序列的对准与评分是Ω结构模式分类的基础。

(1) 序列对准

对准原则:以Dyck树结构的叶子为对准要求。因为生物功能区比较多地集中在结构树的叶子附近,有了对准原则我们可以更方便地考虑两个或多个序列的距离的评分。

(2) 评分准则

评分准则是复杂而多角度的[6],可以认为对于已有文法模式的语言符号串,编辑变换仍然是比较好的基础工具。我们考虑Ω模式的文法串(Ω,·,*)∈Σ对任何两个RNA二级结构A和B,序列A,B∈Σ^*施行三种编辑校正变换T:$\Sigma^*\to\Sigma^*$:

① 替换变换T_s

$$\Omega\cdot\to\Omega *\qquad \Omega *\to\Omega\cdot;$$

② 删除变换T_D

$$\Omega\cdot\to\Phi\qquad \Omega *\to\Phi;$$

③ 插入变换T_I

$$\Phi\to\Omega\cdot\qquad \Phi\to\Omega *$$

使得B=T(A)或相反。这里Φ是空串。

这种变换在研究一维序列进化与句法模式识别时是经常使用的。我们注意Aho-Peterson的观点[7]:两个符号串之间的距离将由一个符号串导出另一个符号串所需要的最小编辑变换的数目来度量。若这个编辑变换只局限于我们上面规定的三种,这就是符号串的Levenshtein度量数[8],它定义为:

$$L_T=d^{\Omega}(x,y)=\min\{\sigma\cdot k_s+\gamma\cdot m_D+\delta\cdot r_1\},$$

k_s,m_D,r_1分别是所需要T_S,T_D,T_I的变换最小数目,不是一个序列二级结构所有的茎、环等在生物功能上有等同重要性,所以导致这三种变换将有不同的权重,这些权重{σ、γ、δ}的研究将是具有生物学内容的。

(3) 拓扑相似度

序列的距离对研究RNA的生物进化特别有意义,作为分类而言,序列的相似度更具有聚类的作用。我们可以有很多方式定义序列的相似度,这里最简单的办法是定义拓扑相似度:

$$L^*(A,B)=1-L(A,B)/(N(A)+N(B))/2,$$

其中 N 为序列碱基数，Levenshtein 数中的权重可以平权，也可以将权重考虑到 L 与归一化因子二者中去。至于怎样做得更具体化，这正是要联系生物实际的问题。

总之，它表征两个 RNA 序列在二级结构上具有相同拓扑结构所占的比例。相似度可以作为分类的一个指标。

(4) 拓扑相似度的细致化

简单的拓扑结构相似一般是粗糙的，近似的，继续引入在相同拓扑结构下一维$\{a,g,c,t\}$序列的同源对准权重作为拓扑相似权重的细微刻画。这在算法上是同工异曲、不费太多力气的。但其意义是将 RNA 一维序列的进化与二级结构有机地结合起来，这是一条新的途径。

A. archaebacterium sulfolbus acidocaldarius

([)(((Ω(((Ω)ΩΩ))((Ω))Ω)ΩΩ)(Ω))ΩΩ(ΩΩ))((((Ω)(((Ω)Ω))))
(Ω)(ΩΩ))(Ω)](((Ω((Ω)((ΩΩ)(ΩΩ)))(((Ω))Ω)Ω)((Ω))Ω

B. cyanobacterium nostcc PCC7120

([)(((Ω(((Ω)ΩΩ)))(Ω)Ω)ΩΩ)(Ω))(Ω)Ω(ΩΩ))((((Ω)(((Ω)Ω))))
(Ω)(ΩΩ))(Ω)](((Ω((ΩΩ)((ΩΩ)((Ω)Ω)))(((Ω))Ω)Ω)(((Ω)))Ω

$L^*(A,B)=0.87$

图 3.3.8　两种小亚基的括弧文本语言与拓扑相似度

4. **生物功能词的分析**

生物功能词的分析，以往只是结合统计词的频数意义去发现词的使用状态。有了 Dyck 文法之后，我们多了一个角度，从语法的推导和结构上去认识这些短符号串的地位、含义，以致可以形式地去分析它的"功能"。这方面的工作除可以类比以往的词分析研究之外，我们认为：(1) 排斥句法的频数统计；(2) 环区与茎区结构统计；(3) 生物功能词的确认，是重要而有意义的工作。

5. **语法分析器**

在模式识别，模式分类，生物词结构研究的基础上可以构造语法分析器。从编译原理讲，语法分析程序是根据文法规则去分析序列中单词组成的各种语法结构的正确性。就产生语法树而言，语法分析方法大致分为自顶向下和自底向上两大类[9]。由于 RNA 的 Dyck 语言在"机器语言"$\{a,g,c,t\}$水平上具有极其复杂的多义性，严格地讲，目前尚未从理论上解决这个难题。我们只能在 Ω 模式给定的基础上，利用自顶而下进行分析。为

此我们创造了三种方法：

① 邦分析方法； ② 团分析方法； ③ 瞬时成型环分析方法。

这些方法主要利用图论的关联与Ω模式的配合而设计。这第三种方法在原则上是模拟真实RNA在形成具有生物学意义的结构中逐段瞬时完成的概念，这对于大型的长序列RNA可能有意义。但这些方法与传统的自由能最小方法是不同的。目前，由于我们掌握的RNA真实结构的标准数据不多，一时很难鉴别优劣。

由语义产生的语法分析器问题，对于分析RNA序列特别重要，因为形式语言并不能干净彻底地解答语言分析中遇到的所有问题，尤其是多义性出现最为突出。传统的语义由RNA序列形成二级结构的环、茎的自由能所担任。如果我们仍然利用这个传统的语义（当然还有其他语义），对于Dyck语言，则需要采取语义的配合上下文无关的厄利算法[10]或者科克-杨格-卡沙米算法。这两种算法目前只停留在原则阶段，估计对于相当长的长链RNA来说，计算复杂性是一个可能遇到的难题。

综上所述，Dyck文法可能是一个分析RNA的可供利用的框架。Dyck文法是否含有生物学功能与意义是最后判定这个理论框架的标准之一，如果我们通过Dyck文法运用自动机的形式语言，不仅在粗粒化与拓扑化上把握住RNA的二级结构，同时通过一级结构的细微化表达RNA的同源与进化特征，则在描述RNA符号序列上可能有重要应用，甚至某些方面将有不可预料的进展。

未来的任务，对于Dyck语法分析RNA而言，是创造性的发展。可用于歧义的上下文无关语言类的算法，特别是适用于Dyck文法的算法，即使对于计算机形式语言的算法发展而言，也是一个重要的尚未完成的任务。

3.3.3 Pseudo-Dyck语言——赝二级结构

目前流行的看法是，RNA大多数二级结构可嵌入平面。Dyck语言将描述这类二级结构中的一大类。RNA结构中还存在其他可嵌入平面结构。例如，比较常见的赝结[11]和开关结构[12]。这些结构在分子生物学研究中已经是为研究者所接受的结构，并也有一定的算法。从语法的角度来认识这类结构是比较复杂的，因为它们已不再属于一般的Dyck语言，其理

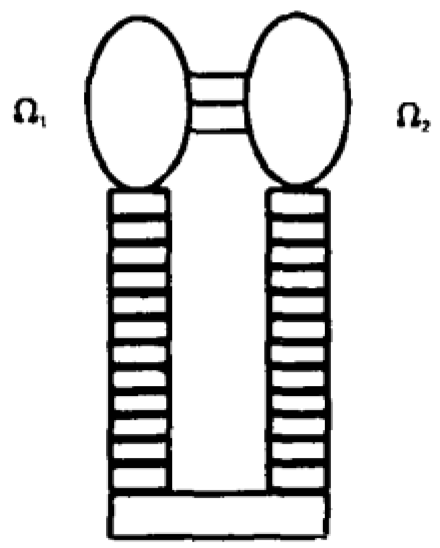
由是：

① 这类语言需要新增加一类句法产生式

$$S \to \omega S \omega^*,$$

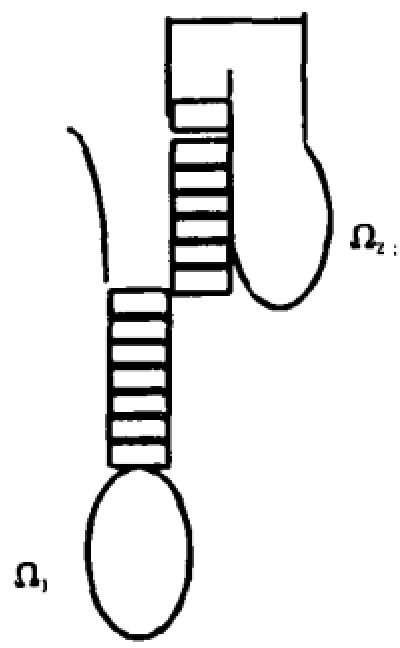
这里 ω^* 是 ω 的沃森-克里克氢键共轭词；

② 由上述增加的句法产生式增加新的代数运算是一种叉乘运算，用以反映 RNA 茎环结构新的几何联接方式(图 3.3.9，图 3.3.10)；



$\Omega_1 \otimes_1 \Omega_2$

图 3.3.9 赝结和赝 Dyck 语言



$\Omega_1 \otimes_2 \Omega_2$

图 3.3.10 纽结和非 Dyck 语言

③ 该种句法产生式将使得 Dyck 语言的括弧文法由单纯的规则嵌套和并列，变成括弧之间的叉花嵌套，即

$$((\ ))(\ )(\ ) \to ([\ )],$$

这样增加了括弧的状态类别，从语言角度这是属于三态 Dyck 语言，已超出上下文无关的阶层而进入上下文有关的范畴。语法代数与语言分析就变得复杂。但是可以肯定这种语言的几何结构仍然是有纽无结(unknoted)的结构，可以称它为赝 Dyck 语言。在 Ω 模式识别时对两种 rRNA 的小亚基结构的分析中，我们已看到这种情况，由于括弧的叉花与类别增加已使 RNA 的二级结构变得多姿多态。所以研究这种语言也是我们重要的任务之一。

3.3.4 Non-Dyck 语言——一种可能的三级结构语言

如果在 Dyck 与平面赝 Dyck 语言的基础上，所有茎区，按其长度转换为纽状的螺旋结构，从而自动地变成三级结构。但这些结构无论

怎样复杂，它们都是有纽无结的。目前在 DNA 双链结构中纽结的研究是相当深入而完美的[13]，纽结理论[14]基本上是可以作为 DNA 分类的一个纯数学标准[15]。但在单链 RNA 结构研究中，虽然有局部的螺旋纽状，一般还未注意 RNA 的纽结问题。这里只单纯从语言角度来探讨这个问题，也是一个相当有趣的事情，我们来简单作些启发式的讨论以丰富我们的 Dyck 语言结构。设想有图 3.3.11 的例子存在，我们就得到了一个具有纽结的 RNA，它也许不存在于现实生物中。但它的语言已超出 Dyck 文法了。在进一步讨论中我们还可以增加一种新的句法产生式：

$$S \to \omega S \omega^R,$$

这里 ω^R 是 ω 词的镜象态，由于镜象态的增加，必须增加第二种、甚至第三种叉乘的代数运算来反映 RNA 茎环结构的又一种新的联接方式（参看图）。同时括弧文法由二类三态叉花乘变成三类四态叉花乘。

(( ))(( ))([)]→([)({])}……，

这时 Dyck 语言集的括弧是广义四态的，它包括象原 ω，厄米共轭态 ω^+，共轭态 ω^* 与镜象态 ω^R 四种状态词。这时 Dyck 语言一定超越乔姆斯基第二阶层进入第三阶层，甚至有可能局部地进入更高阶层语言。我们可以称为非 Dyck 语言。纽结与非 Dyck 语言的关系复杂，尚需仔细鉴定。

若在现实中，正如我们理论所想象的那样，RNA 的三级结构产生了

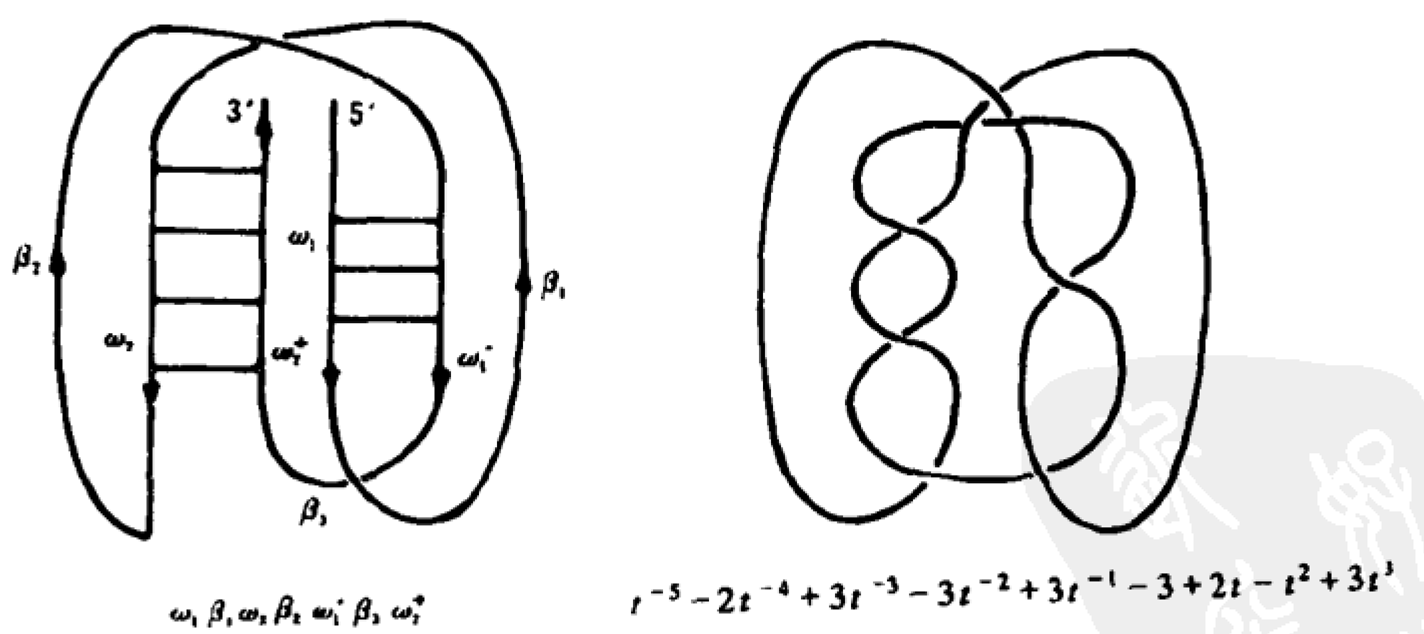


图 3.3.11　猜想的 RNA 纽结与 Jones 多项式

真正的纽结(knot)结构,那么熟知的 Jones 多项式可以作为描述 RNA 结构的一种有用拓扑代数工具之一。这类工具的使用,就可以将 RNA 序列的描述引导到更为优美的数学境地。我们可以猜想,生物 RNA 序列的结构在多大程度上喜欢纽结结构。如果一旦喜欢成为普遍的倾向,那么可以提出一个重要问题:给定($\omega_1\omega_1^+\cdots\omega_n\omega_n^+$)集中一个序列,它在空间排布成最小幂 Jones 多项式的几何结构是什么?如果生物遵从这个原理,似乎有这种可能性去找到一种关于 RNA 三级结构的解答,但不知道客观世界是否像允许力学质点走最短的测地线一样,同样允许我们对生物提出最小 Jones 多项式所对应的结构问题。

参考文献

[1] Chomsky N. (1956) PGIT 2: 3 *Three models for the description of language*, 1959, *On certain formal propertise of rammers*, 1959, *Inf and Control*, **2**: 137

[2] 彭守礼,刘次全. 论基因信息结构复杂性. 第24次香山会议报告,1994, 11

[3] 刘次全. 量子生物学及其应用. 北京:高等教育出版社, 1992

[4] Searls D B. *American Scientist*, 1992, **80**: 579

[5] Hoporotf J E, Uuman J D. *Formal Languages and Their Relation to Automata*. 1969; Addison-Wesley Publishing Co. Reading Mass. Harrison, M. Introduction to Language Theory. Addison-Wesley 1978

[6] Vingron M, Waterman M S. *J Mol Biol*, 1994, **235**: 1

[7] Aho A V, Peterson T G. *SIAM J Comput*, 1992, **1**: 305

[8] Fu K S(ed). Syntactic Pattern Recognition, Application. Springer, 1977

[9] 蒋立源. 编译原理. 西北工业大学出版社, 1994

[10] Kasame T, Torii K. *JACM*, **16**(1069): 423

[11] Puglisi J D, Wyatt J R, Jinoco Jr I. *Nature*, 1988, **331**: 21

[12] Fasman G D. *Handbook of Biochemistry and Molecular Biology*, *vol* Ⅱ *Nucleic Acids*, CRC Press, 1976, 328

[13] Wang J C. *Annu Rev Biochem*, 1985, **54**: 665

[14] Kauffman L H. *On Knots*. Princeton University Press, 1987

[15] White J H, Millett K C, Coggarelli N R. *J Mol Biol*, 1987, **197**: 585

(彭守礼 刘次全)

§3.4 基因组序列的大尺度同源分析

国际基因组计划的主要目的是要弄清所有基因蛋白质的功能。至

今在基因库中存放的 DNA 序列已超过 10^9 个碱基对,其中大多数基因蛋白质的功能尚未确定。而且,这种积累势头还在继续,几乎每两年翻一番。幸好,DNA 序列的同源分析法可以确定大约一半的基因蛋白质的功能。目前有两类流行的同源分析软件:一类以 BLAST[1] 为代表,采用快速的字匹配搜索方法;另一类以 SIM[2] 为代表,采用严格的动态规划算法。它们都很难用于基因组序列的大尺度同源分析,因为 BLAST 显示大量无间隔(即连续)的匹配片段后,用户需设法把它们连接成具有生物学意义的序列联配(局部的或整体的),这在基因组的尺度上,需要浪费很多时间去排除大量实际上无意义的匹配片段。而 SIM 用于基因组分析,所花的时间太长。因此这两类方法均不适用于大尺度基因组序列分析。这里介绍一种足够精确且十分快捷的大尺度同源分析方法。经植物叶绿体基因组,哺乳类 T 细胞受体基因组和哺乳类眼 γ 晶状体基因簇等实例应用,证实此方法达到了速度和精度间的良性平衡,可用于实际的基因组序列分析。

3.4.1 大尺度同源分析方法

设两 DNA 序列为 $A=A_1A_2\cdots A_m$ 与 $B=B_1B_2\cdots B_n$,它们的大尺度同源分析的整个算法程序分三个步骤,可用图 3.4.1 表示。

第一步,利用散列技术快速检索长度至少为 l 的匹配片段,见图 3.4.1(A)。

用 0,1,2,3 编码碱基 A,G,C,T,并计算 A 序列中所有长度为 l 的片段(在位置 i)的四进制码,并将之翻译成十进制码。按片段代码将这些片段变换成散列数据结构,即可以按照代码快速查找片段在 A 序列中的位置 i。同样计算 B 序列中所有 l 片段(在位置 j)的代码,按此代码找到 A 序列中具有此代码的片段(可能不止一条),并记录下找到的匹配片段 (i,j)。检查所有同一对角线上的匹配片段,即 $(j-i)$ 在此线上不变,如果有两条匹配片段 (i,j),(i',j') 相重,即若 $i-l\geqslant i'$,则将这两条匹配片段合并成一条且记为 $(i,j,l+(i'-i))$,表示起点在 (i,j) 的长度为 $l+(i'-i)$ 的匹配片段。如果起点在 (i,j),长度为 l 的匹配片段与同一对角线上的任一条片段均不相重,则记此片段为 (i,j,l)。在记录匹配片段时按照对角数 $(j-i)$ 分类,属同一对角数的再按 i 的大小排序。这一步的操作很快完成。

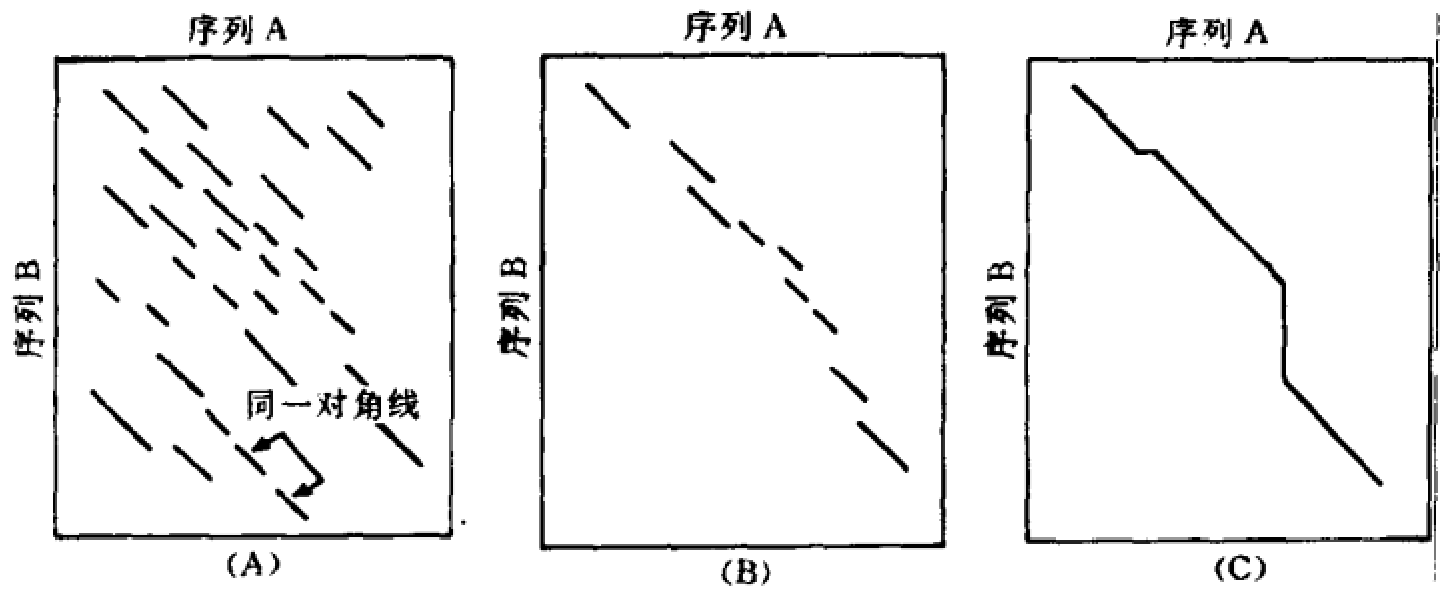


图 3.4.1 本文所述计算方法的打点法表示

(A) 用散列技术搜索两序列中无空缺的匹配片段。其中每一条短线代表一个匹配片段。(B) 将搜索到的片段集合优化成一个合乎生物学意义的片段子集。(C) 将子集中的片段连接成长的联配。

第二步，把前一步获得的大量匹配片段优化精简，以便从中找到与期望的序列联配最关联的匹配片段子集，见图 3.4.1(B)。用经我们改进的 Eppestein 等人[3]的稀疏动态规划法可以做到这点。此算法的递推格式为：

$$D(i,j,k)=\mathrm{sc}(i,j,k)+\max\{L(i,j,k),M(i,j,k),R(i,j,k),0\}, \tag{3.4.1}$$

$$L(i,j,k)=\max_{\substack{i'<i,j'<j\\ j'-i'<j-i}}\{D(i',j',k')+\mathrm{gap}((j-i)-(j'-i'))+\mathrm{mut}(i-i'-k')\}, \tag{3.4.2}$$

$$M(i,j,k)=\max_{\substack{i'<i,j'<j\\ j'-i'=j-i}}\{D(i',j',k')+\mathrm{mut}(j-j'-k')\}, \tag{3.4.3}$$

$$R(i,j,k)=\max_{\substack{i'<i,j'<j\\ j'-i'>j-i}}\{D(i',j',k')+\mathrm{gap}((j'-i')-(j-i))+\mathrm{mut}(j-j'-k')\}, \tag{3.4.4}$$

其中 $\mathrm{gap}(t)=g+t\cdot e$ 是刻划删去或插入突变的间隙罚分函数，而

$$\mathrm{mut}(t)=\begin{cases}t\cdot r, & \text{当 } t>0\\ t\cdot s, & \text{当 } t\leqslant 0\end{cases}$$

是刻划取代突变的罚分函数。sc (i, j, k) 是匹配片段 (i, j, k) 上

的保守性分数。一般取 $s=20$，$r=-5$，$g=-3$，$e=-2$。如图 3.4.2 所示，从 (i', j', k') 到 (i, j, k) 只有三种路径，沿对角线上 M，沿对角线上方 R，沿对角线下方 L。比如沿对角线上方 R，则最佳路径应是粗黑线。所以，到 (i,j,k) 为止的总分数包括：此片段上的完全保守性分数 $sc(i,j,k)$，到片段 R 为止的最佳联配分数 $D(i,j',k')$，扣掉在 $D(i',j',k')$ 中多算的一段（即 R 片段中额定细线部分）的保守性分数，即(3.4.4)中的 mut 部分，以及粗黑线的垂直部分对应的插入突变罚分。再经极值化即得(3.4.4)式。类似地可理解(3.4.2)和(3.4.3)。最后从三种路径中取最高联配分数作为 $D(i,j,k)$，得(3.4.1)式。

按 i,j 从小到大的次序对每一条匹配片段 (i,j,k) 均计算一次 $D(i,j,k)$，然后记录下与此匹配片段联络的上游匹配片段 (i',j',k')。如果算得的最佳分数仅为其自身的保守性分数 $sc(i,j,k)$，则记 (i,j,k) 为一条路径的起点。当所有的最佳联配分数算得后，从其中分数最大的一个片段出发回溯，就可以找到一个最佳的匹配片段集合。换言之，这个优化集合虽比原匹配片段的集合小得多，但却保存了有关两个序列之间的同源信息。

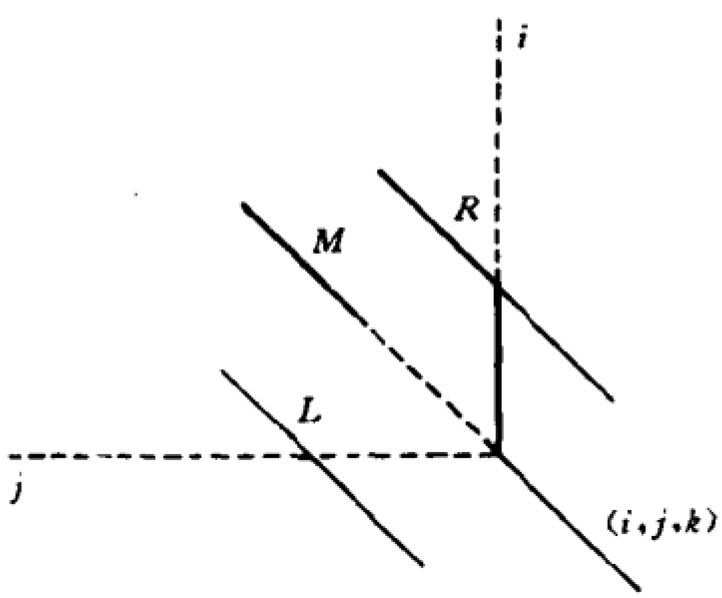
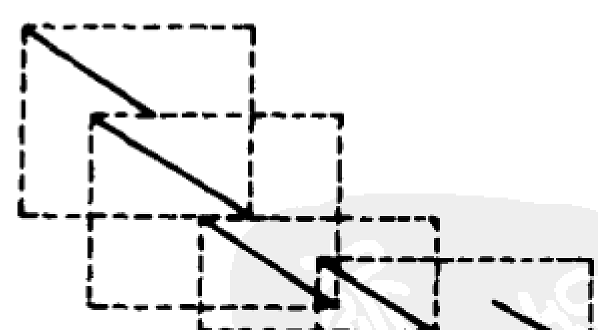


图 3.4.2 稀疏动态规划的具体过程


图 3.4.3 由片段链获得约束动态规划所需的边界

第三步，前面得到了最佳联配片段集，但是还没有连接。因此，为了获得两序列间完整的同源信息，必须实现片段连接。设前步找到的片段集合为：$(i_1,j_1,k_1),(i_2,j_2,k_2),\cdots,(i_s,j_s,k_s)$，其中 s 为片段总数。则如图 3.4.3

所示，从起点(i_1,j_1)到终点(i_s,j_s)的狭窄区域内再次利用动态规划实现序列联配，迭代格式为：

$$M_{ij}=\max\{M_{i-1,j-1}+S_{ij},F_{ij},G_{ij}\}, \tag{3.4.5}$$

$$F_{ij}=\max\{M_{i-1,j}+g,F_{i-1,j}\}+e, \tag{3.4.6}$$

$$G_{ij}=\max\{M_{i,j-1}+g,G_{i,j-1}\}-e, \tag{3.4.7}$$

$$S_{ij}=\begin{cases} s, & 当 \quad A_i=B_j \\ r, & 当 \quad A_i\neq B_j, \end{cases} \tag{3.4.8}$$

M_{ij} 是 $A_{i1}A_{i1+1}\cdots A_i$ 和 $B_{j1}B_{j1+1}\cdots B_j$ 的最佳联配分数。

而最佳匹配片段集可用来确定此区域的左右边界，即每一片段的头与尾随片段的末尾确定一个矩形，由这些矩形形成的上包络和下包络来规定动态规划的区域。与此类似的联配方案也有人用过[4]。此时，参数应调整为 $s=3,r=-1,g=-3,e=-1$。

3.4.2 同源分析方法的应用

按照如上计算方法，我们制作了软件 LALIGN(取片段长度 $l=8$)，并对植物叶绿体基因组，哺乳类 T 细胞受体基因组 C_α/C_δ 区域和哺乳类眼 γ-晶状体基因簇等三个实例进行了联配，结果如下。

1. 植物叶绿体基因组

植物叶绿体基因组的全序列测定完成较早。烟草叶绿体基因组(长度 121 024 bp)与苔类植物地钱叶绿体基因组(长度 155 844 bp)的联配是一个很具挑战性的问题。从生物学的角度来看，它们分别代表了植物的两种主要亚类：维管类植物(烟草)和非维管类植物(地钱)。它们被认为是在 4 亿年前由共同的祖先分化来的。尽管如此，两类植物叶绿体基因组仍保持了相近的基因组织情况[5]。

两基因组经 LALIGN 联配后用打点法表示成图 3.4.4。较长的一个联配是两基因组正链联配结果。较短的一个是烟草叶绿体基因组的正链与地钱叶绿体基因组的互补链联配的结果，表示地钱叶绿体基因组上从 *trnH* 到 *trnC* 的多个基因处在它的互补链上。由此可得两基因组的基因匹配情况，见表 3.4.1。可见，两基因组在外显子水平上保持了相当高的同源性。所用机时约为 3 个小时(在SGI R3000工作站上)，与此对比，SIM 的机时[4]约为一个星期。

表 3.4.1　由烟草叶绿体基因组与地钱叶绿体基因组联配得到的基因序列相似性

对应较长联配结果的外显子序列相似性(平均为 52.9%)					
地钱叶绿体基因组上的基因名	烟草叶绿体基因组上的基因名	相似性(%)	地钱叶绿体基因组上的基因名	烟草叶绿体基因组上的基因名	相似性(%)
trnD	*trnD*	93.2	*trnW*	*trnW*	
trnY	*trnY*	90.4	*trnP*	*trnP*	93.2
trnE	*trnE*	91.4	*rp*133	*rp*133	72.2
trnT	*trnT*	95.4	*rps*18	*rps*18	69.2
psbD	*psbD*	84.3	*rp*120	*rp*120	65
psbC	*psbC*	84.8	*psbB*	*psbB*	84.2
trnS	*trnS*	83.6	*rpoA*	*rpoA*	68.3
*ORF*62	*ORF*105	82.5	*rps*11	*rps*11	74.8
trnG	*trnG*	87.3	*infA*	*infA*	70.4
trnM	*trnM*	91.8	*rps*8	*rps*8	67.8
*rps*14	*rps*14	72.4	*rp*114	*rp*114	79.6
psaB	*psaB*	84	*rp*116	*rp*116	75.2
psaA	*psaA*	83.8	*rps*3	*rps*3	66
trnS	*trnS*	83.3	*rps*19	*rps*19	82.3
*rps*4	*rps*4	74.8	*rp*12	*rp*12	74.2
trnT	*trnT*	83.5	*rp*123	*rp*123	68.9
trnL	*trnL*	55.3	*trnI*	*trnI*	93.2
trnF	*trnF*	74.9	16 *S RNA*	16 *S RNA*	91.5
*ORF*169	*ORF*168	72.4	23 *S RNA*	23 *S RNA*	93.8
psbG、*ndh*3	*bhp*	53.4	4.5 *S RNA*	4.5 *S RNA*	90.2
trnV	*trnV*	89.1	5 *S RNA*	5 *S RNA*	92.2
trnM	*trnM*	97.3	*trnR*	*trnR*	98.4
psbF	*psbF*	85.9	*trnN*	*trnN*	94.2
psbE	*psbE*	83.3	*rps*15	*rps*15	66.6
对应较长联配结果的外显子序列相似性(平均为 56.6%)					
trnH	*trnH*	92	*atpA*	*atpA*	80.7
psbA	*psbA*	87.5	*atpF*	*atpF*	67.5
trnK	*trnK*	88.5	*atpH*	*atpH*	85.4
trnQ	*trnQ*	65.1	*atpI*	*atpI*	74.3
trnS	*trnS*	89.7	*rps*2	*rps*2	76.2
trnG	*trnG*	63.3	*rpoB*	*rpoB*	74
trnR	*trnR*	86.6	*trnC*	*trnC*	93.6

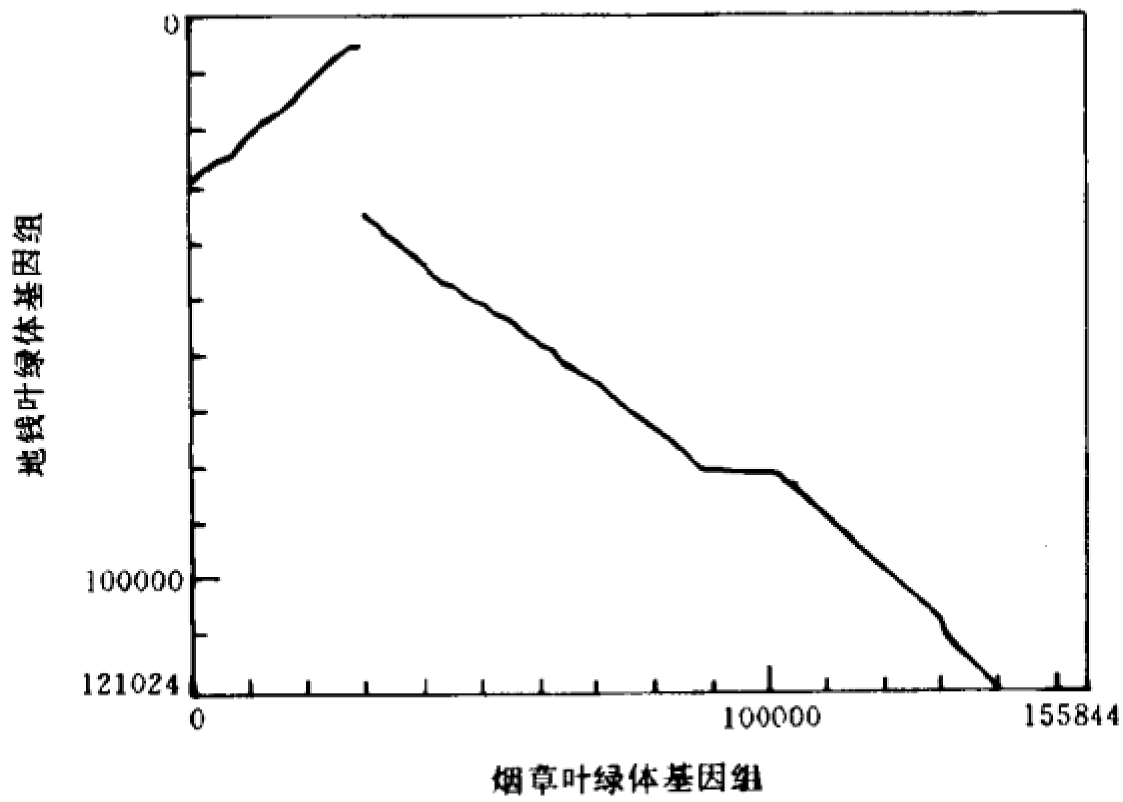
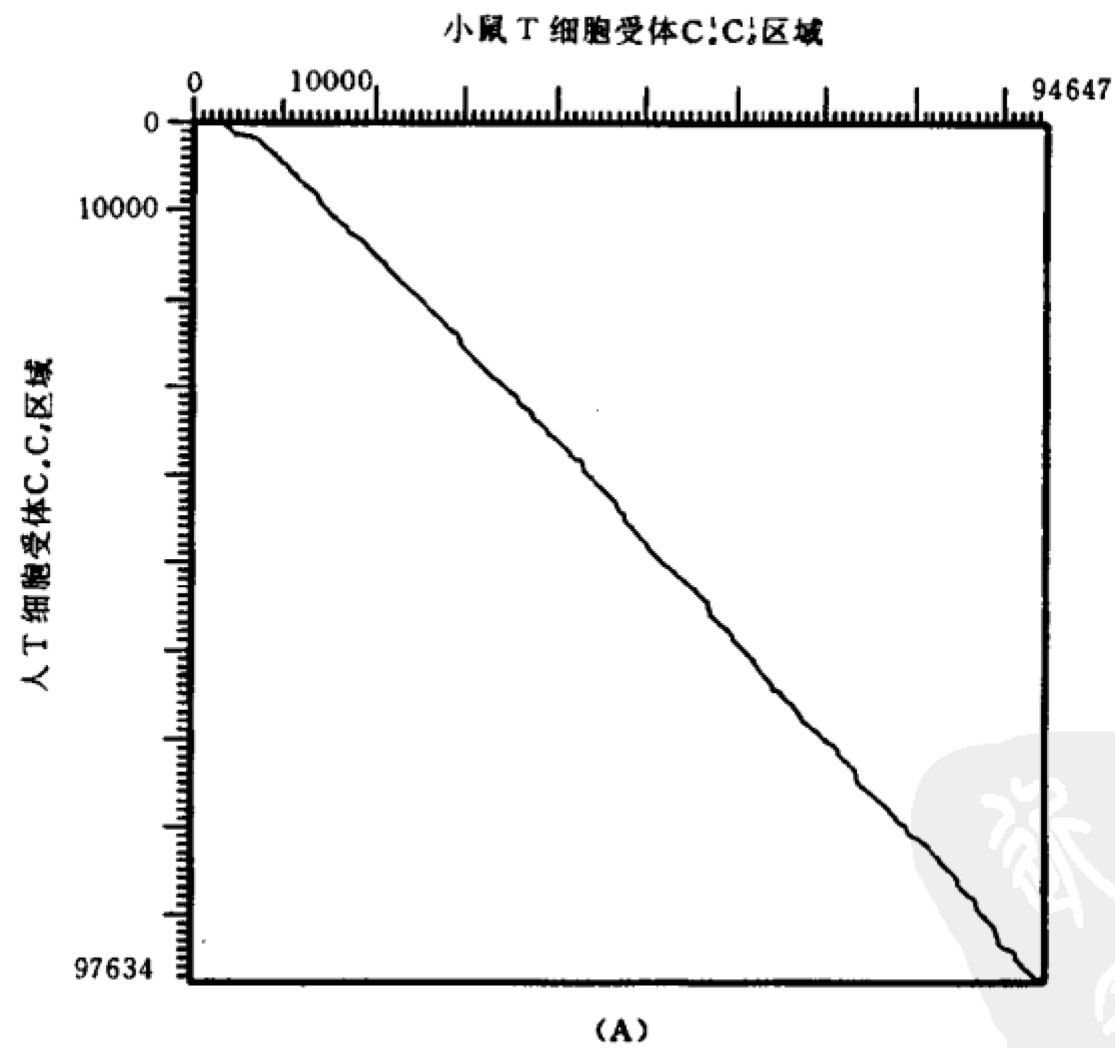


图 3.4.4 烟草叶绿体基因组与地钱叶绿体基因组联配的打点法表示

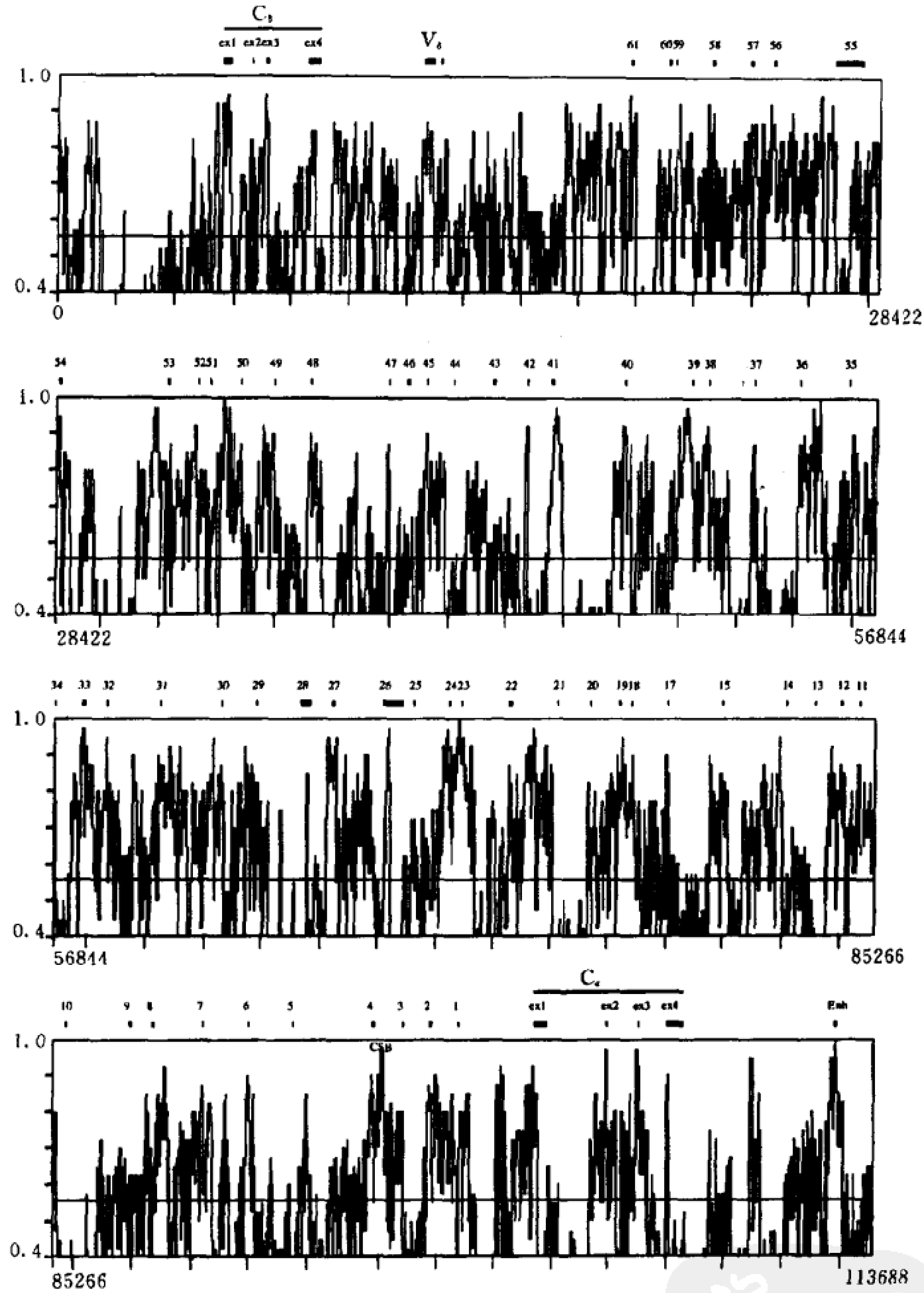
2. 哺乳类 T 细胞基因组 C_α/C_δ 区域

T 细胞在免疫应答反应中扮演了一个重要的角色。T 细胞受体能够识别外源大分子(抗原),然后促进其他淋巴细胞的分化。另一方面,它又



(A)

图 3.4.5



(B)

图 3.4.5 小鼠与人的 T 细胞受体 C_α/C_δ 区域的打点图(A)和扫描分析图(B)

其中 E_{nh} 表示 3′增强子；1～61 表示 J_α 基因外显子

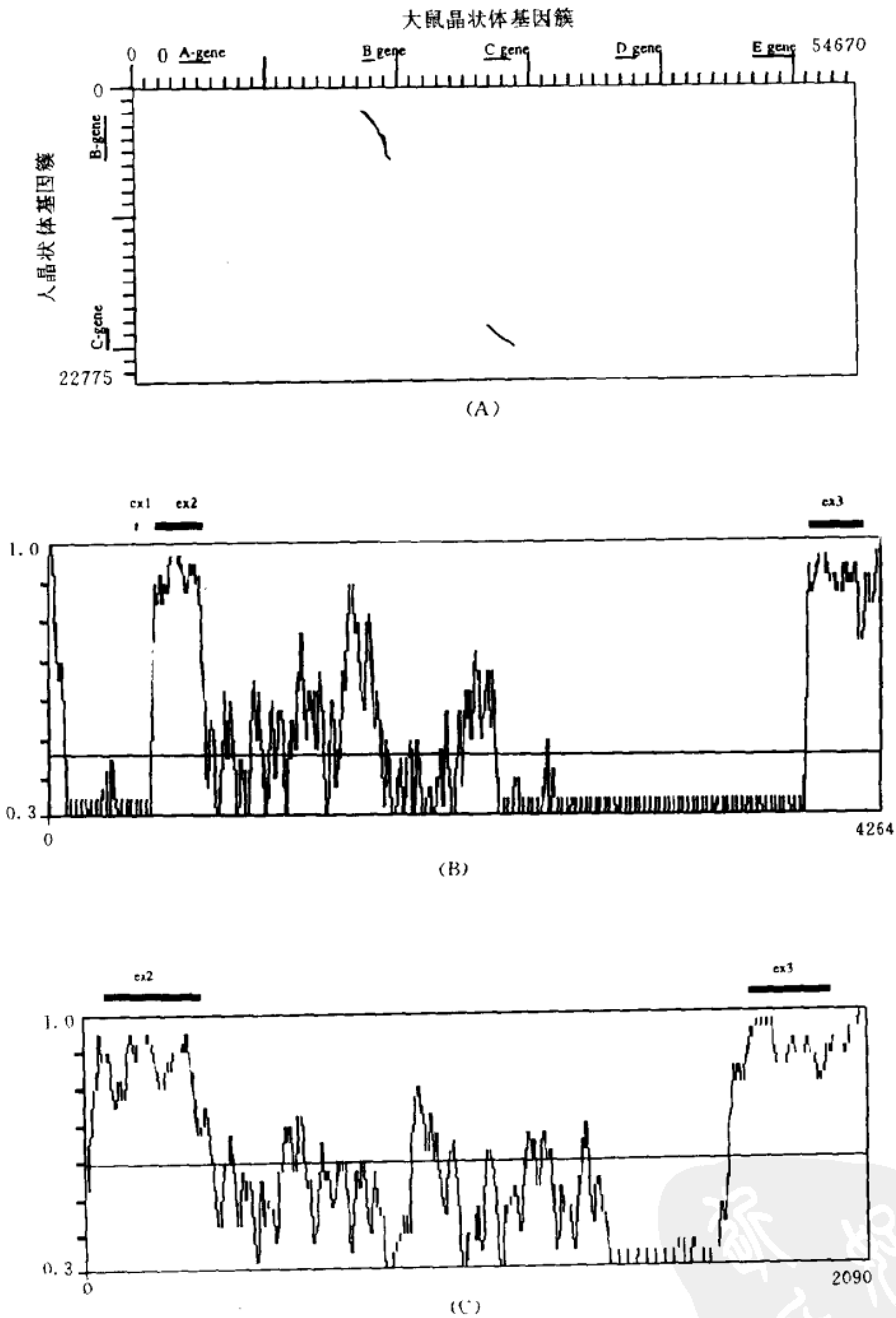


图 3.4.6　大鼠与人的眼 γ 晶状体基因簇联配的打点图(A)和扫描分析图(B,C)

可以识别组织相容性抗原(MHC),这种双重识别功能是免疫应答反应中的重要一环[6]。

用 LALIGN 把它们联配,且用打点法和扫描分析法表示成图3.4.5。从中可以看出:整个区域的相似性很高(约 56%),而编码区在其中仅占很少比例(约 8%),这表明在这一区域非编码区也相当保守。3′增强子有较高保守性。另外,在 J_α 区域内含子中存在一个高度保守的片段(简称为CSB),长度约150 bp,保守性高达 95%以上,可能存在一定的功能。这些结论与 Hood 等人的结果相符[7]。

3. 哺乳类眼 γ 晶状体基因簇

与 T 细胞受体的例子形成对照,人和小鼠眼 γ-晶状体基因簇的外显子比内含子具有高得多的同源性。经用 LALIGN 分析,它们之间的整体同源性较低,但两个局部同源区是明显的,见图 3.4.6(A)。在两个局部同源区作扫描分析(图 3.4.6(B)和(C)),可以看到外显子有较高的同源性(约 90%),而内含子的同源性较低。Dunnen 等人在这两个家族基因组中均发现大量重复序列,并认为这是在物种基因分叉后插入的[8],这也可能是二者相似性较低的一个原因。

从上面三例结果可以证实,LALIGN 是基因组序列大尺度同源分析的便捷工具。更重要的是,这种片段扫描、优化、再连接的方法可能是大尺度基因组分析的一个较为理想的近似方法,可以进一步推广到基因组分析的其他应用中。

参 考 文 献

[1] Altschul S, *et al.* *J Mol Biol*, 1990, **215**: 403
[2] Huang X, Miller W. *Adv Appl Math*, 1991, **50**: 97
[3] Eppestein D, *et al.* *J Assoc Comput Math*, 1992, **39**: 519
[4] Chao K, *et al.* *Bull Math Biol*, 1993, **55**: 503
[5] Palmer J. The molecular biology of plasmids. *Vol* 7. Cell culture and somatic cell genetics in plants. Springer Press, 1984
[6] Jorgenson J, *et al.* *Annu Rev Immunol*, 1992, **10**: 835
[7] Hood L, *et al.* In: Genomic analysis. Vol 5. Regional physical mapping. Cold spring harbor laborotory press, 1993
[8] Dunnen J T, *et al.* *Gene*, 1989, **78**: 201

(汤海旭 丁达夫)

§3.5 符号序列的复杂性刻画

符号序列的复杂性分析在理论及应用上都是非常有意义的研究课题，在这方面已有很多讨论。本节的目的是介绍这个方向上的一些进展，其中包括苏州大学非线性中心所做的工作。

3.5.1 介绍柯尔莫戈洛夫(Kolmogorov)复杂性的基本概念。这是较早提出的关于单个有限符号序列的一种复杂性刻画方法，已有多方面的应用。同时也指出其不足之处和有关文献。

3.5.2 介绍动力系统中的复杂性研究概况。这方面的发展有可能导致新的科学分支形成。这里我们着重指出，形式语言和自动机等工具在符号序列的复杂性分析中将起重要作用。

最后一小节讨论几个专题。其看法尚不成熟，只是希望起到抛砖引玉的作用。其中有的是我们自己的工作，有的是对某些问题的概述。

由于不可能在一篇短文中将许多问题说清楚，本节列举出较多的文献以补充不足，希望得到读者谅解。

3.5.1 柯尔莫戈洛夫复杂性

在 60 年代中期柯尔莫戈洛夫，Chaitin 和 Solomonoff 分别独立地提出了刻画符号序列复杂性的相同方法。在这种方法中，对于给定的一个有限长度的符号序列 x，将能够生成这个序列的最短计算机程序 p 的长度 $|p|$ 作为序列的复杂性度量。为了摆脱与具体计算机的依赖性，这里的计算机是指通用图灵机。我们称 $|p|$ 为符号序列 x 的柯尔莫戈洛夫复杂性，记为 $K(x)=|p|$。

举例来说，对于一个重复多次的周期序列

$$x=(10)^n=101010\cdots10,$$

只需要用含义如下的简单程序“p= 重复 n 次打印符号 10”就可以生成 x。将这个程序用二进制代码写出来，则当 n 充分大时，程序 p 的长度为

$$|p|=\log_2 n+c,$$

其中 c 是一个与 n 无关的常数。由此可见，当 n 充分大时，序列 x 的柯尔莫戈洛夫复杂性 $K(x)$ 远小于 x 的长度 $|x|$。这也就是说 x 是可以大大压缩的。我们称这样的符号序列其复杂性是很低的。反之，如果 $K(x)$ 与 $|x|$

相差无几，则认为 x 的复杂性是很高的。

再举一个例子，考虑圆周率 π 的近似值，它在目前已计算到五百亿位以上，但仍然看不出有什么规律性。用数学的语言来说，这些近似值能通过所有的随机数检验，因此从统计角度来看它是随机序列。但是这些位数非常长的近似值可以用相对来说很短的程序来生成，因此它的柯尔莫戈洛夫复杂性是很低的。

柯尔莫戈洛夫复杂性是一个严格的数学理论，并有多方面的应用。对此有兴趣的读者可以参看新近出版的专著[1]，一个简单的介绍见文献[2]的第七章。

在柯尔莫戈洛夫复杂性的众多性质中，下列几点是引人注目的。

(1) 在一定的严格数学意义上，柯尔莫戈洛夫复杂性是不可计算的。

(2) 长度充分大的序列中，绝大多数的柯尔莫戈洛夫复杂性都很高，即不能压缩。

这里的第一点由于出现了在文献[3]中的算法而得到一定程度的解决(见文献[4])。当然这时所计算的并不是原来的 $K(x)$ 了。我们更关心的是第二点的实际意义。事实上它就是说绝大多数序列都像是随机序列，杂乱无章而不能压缩。这同我们的日常经验完全一致。但这也表明，柯尔莫戈洛夫复杂性很难帮助我们寻找隐藏在符号序列中的规律性。

3.5.2 动力系统中的复杂性研究

近几十年以来，在非线性科学方面的一系列发现使人们认识到复杂现象可以从简单系统中产生，几个典型例子如下：

(1) 含单参数的单峰区间映射族，其中又以抛物线映射(也称 Logistic 映射)

$$x_{n+1}=\mu x_n(1-x_n),\quad 0\leqslant x\leqslant 1,\quad 0<\mu\leqslant 4,$$

最为著名。这个迭代公式极其简单，但当参数 μ 从 3 变到 4 时，由迭代所得的轨线行为非常复杂，而且具有普适性。这方面的文献很多，例如[5～9]等，其中生物学家梅(May)在 1976 年发表于 *Nature* 上的文章[8]有很大的影响。还应指出，抛物线映射即是在生物学中由 Verhulst 于 1845 年(见 F. Cramer 的著作《混沌与秩序，生命的复杂结构》)提出的生长规律。

(2) 通过在复数域中的二次映射

$$z_{n+1}=z_n^2+c,$$

可以生成非常复杂而美丽的曼德布罗特(Mandelbrot)集。这方面的内容可以参看[10]和[11,12]中的讨论。

(3) 洛伦兹(Lorenz)在60年代初对天气模型的研究中得到常微分方程组

$$\dot{x}=\sigma(y-z),\ \dot{y}=(r-z)x-y,\ \dot{z}=xy-bz,$$

其中σ,r和b是参数。虽然在方程组的右方只不过出现了两个二次项,但它却是形成奇怪吸引子概念的第一个例子,由此而命名的蝴蝶效应改变了人们对长期天气预报的许多看法[11,13,14]。

(4) 在冯·诺依曼(von Neumann)自繁殖机研究的基础上,Conway于70年代初提出了著名的生命游戏。这实际上是一种二维元胞自动机,由非常简单的规则可以产生极为复杂和难以预测的行为[15]。

(5) 日本学者上田完亮(Yoshisuke Ueda)在1961年用计算机模拟发现了非线性周期受迫振子(Duffing方程)中的混沌现象,但当时未得到承认,直到1978年才为国外所知,并被称为日本吸引子[16]。

随着非线性科学的发展,开始了以复杂性本身为研究对象的探讨,这方面的研究机构也纷纷出现。其实早在50年代,冯·诺依曼就已经预测到复杂性研究将是20世纪科学的任务。以下是从文献[17]中转引的一段文字:

> "本世纪50年代冯·诺依曼预言,20世纪科学家面临的任务是阐明复杂性(complexity and complication)的概念,正像19世纪的科学面对能量和熵的概念一样。"

但从目前关于复杂性和复杂系统的研究情况来看,还只能说处于初创阶段,这方面可参看文献[18,19]和文献[20,21]等应用例子。有的科学家预测,这方面的研究工作有可能在21世纪发展成为一门或几门科学(见文献[22]的引言)。

这里需要说明,本节所说的复杂性研究同计算的复杂性理论不同。前者认为复杂性是介于规则性和随机性之间的未知的研究领域,而后者以分析算法的时间和空间复杂性为研究对象,其中最重要的未解决问题是P和NP是否相等[12,23,24]。当然,在方法和工具上两类研究是可以相通的。

由于作者所知甚少,本节的以下部分只对动力系统研究中讨论较多

的符号序列复杂性分析作一个简要介绍。这里的多数工作是以上面提到的单峰区间映射和一维元胞自动机作为对象来进行讨论的。因为对于前者已有比较清楚的了解，从而在粗粒化之后可以提供许多性态已知的符号序列[6,9,25,26]，而后者则很容易在微型计算机上进行实验。

将柯尔莫戈洛夫复杂性理论用于动力系统的早期工作见文献[27]。从文献[28]可以知道，至少对于单峰区间映射来说，柯尔莫戈洛夫复杂性同李雅普洛夫(Lyapunov)指数、测度熵等刻画手段基本上是一致的。这并不奇怪，因为对动力系统来说，柯尔莫戈洛夫复杂性主要是度量序列的随机性，而随机性不等于复杂性。

通过 Wolfram，Grassberger 等人的工作[29~33]，计算机科学理论中的形式语言和自动机开始成为动力系统的复杂性研究中的重要工具。这方面可以参见文献[24，33～35]等综述文章。拙作[2]是非线性科学丛书中的一种，其中反映了到 1993 年初为止的研究状况。

事实上形式语言和自动机是离散数学中比较成熟的部分。如果将上述发展看成是这些工具在动力系统研究中的渗透，则还可以举出在模式识别领域中的形式语言方法和在控制论领域中新近发展起来的离散事件系统作为相同趋势的例子。因此我们可以认为，形式语言和自动机在动力系统中的应用是非常自然的，它是符号动力学的一个发展。

3.5.3 对几个问题的看法

在这一节中进一步介绍在复杂性分析中的几个专门问题。

1. D 类语言与禁止字

容易理解，在动力系统中所出现的形式语言一定会具有自己的特点。吸取在文献[29，36，37]等工作中的思想，我们在文献[2，35，38]中提出了 D 类语言的概念，这里的 D 是 Dynamical Systems 的第一个字母。设 S 是一个由有限个符号组成的集合，即所谓字母表。用记号 S^* 表示由 S 中的符号构成的所有有限长度的符号序列全体的集合。所谓形式语言就是 S^* 的任何一个子集合。称一个形式语言 $L\subset S^*$ 为 D 类语言，如果它满足以下两个条件：

(1) 如果 $x\in L$，y 是 x 的子串，则 $y\in L$，

(2) 如果 $x\in L$，则存在 $a\in S$，使 $xa\in L$。

这两个条件的动力学意义是明显的。

对于D类语言来说，在文献[29]等工作中出现的禁止字概念是很有用的。称 $x \notin L$ 为 L 的禁止字，如果 x 的每一个真子串都属于 L。如果记语言在 S^* 中的补集为 L'，记 L 的禁止字全体的集合为 L''，则对D类语言成立公式

$$L' = S^* L'' S^*, \quad L = S^* - S^* L'' S^*。$$

因此禁止字是对语言的一个完整描述。举例来说，当单峰区间映射有稳定的周期3轨道时，我们知道同时存在具有任何周期的周期轨道，且具有Li-Yorke意义下的混沌(见文献[39])。如果这时的揉序列为$(101)^\infty$，则可以证明相应的语言只有唯一的一个禁止字100，它决定了这时的全部符号动力学行为。

关于禁止字的讨论见文献[2]的第五章。关于 L 和 L'' 在乔姆斯基层次中的对应关系的讨论见文献[38]。

2. 乔姆斯基层次体系不是唯一的选择

在讨论语言复杂性时最常用的方法就是用乔姆斯基的层次体系来作为比较各种语言复杂程度的依据。乔姆斯基体系是一种生成语法系统，其中有4个层次，按照其语法的复杂程度从低到高的顺序为正规语言、上下文无关语言、上下文有关语言和递归可枚举语言。它们分别是有限自动机、下推自动机、线性界限自动机和图灵机所能识别的语言。对于正规语言来说，各种计算都是比较容易的。我们经常将接受一个正规语言的最小确定性有限自动机的状态个数作为这个语言的复杂性的度量。但是对上下文无关语言来说，就没有这样方便了。这里出现了不可判定问题，即不存在任何算法来处理的问题。对于上下文有关语言则有更大的困难，要对一个简单例子写出它的语法规则都不是直接了当的事。而为了举出一个例子，使它是递归可枚举语言，但不是上下文有关语言，则必须使用康托(Cantor)的对角线方法，即非构造性的方法(以上可参看文献[23,24]等)。

因此在文献[33]中提出了考虑其他语言系统的想法，例如由生物学家Lindenmayer在研究生物发育模型时提出的一种并行重写系统(即 L 系统)，就有应用的可能性，具体的例子见文献[41,42](关于 L 系统的知识可以参看文献[40])。

3. 对否定性结果的看法

在用计算机科学中的工具来讨论问题时，往往不可避免会与否定性结果打交道。除了上面已提到的不可判定性，还有不可计算性，难解性，哥德尔(Gödel)不完全性等，例子见文献[12,43]和杂志 *Chaos*, *Solitons and Fractals* 的 1995 年第 2 期(这是这方面的一期专辑)。事实上，由于图灵机只有可列个，因此在有不可列个计算对象时就一定会出现不可计算的复杂程度。例如在最简单的一维单峰区间映射和圆同胚映射的情况就是如此。更简单的例子就是实数系。由于同样的理由，在所有实数中只有可列个是可计算的。

应当指出，上述否定性结果并非是消极的东西。它们揭示了客观上必须遵循的规律，具有极为深刻的理论意义(参看文献[12])。另一方面，再以实数系为例，仅仅只有可列个实数是可计算的这一结论，丝毫也没有妨碍科学家对实数的使用。我们相信，形式语言和自动机等工具在复杂性分析中的应用，并不会由于否定性结果的存在而举步不前。

4. 具有无限个状态的自动机

作为一个尝试，对于非正规语言的情况，我们直接考虑具有无限个状态的无限自动机的普适结构，并在单峰区间映射和圆同胚映射上取得了成功。

应当指出，在文献[31]中作出了费根鲍姆(Feigenbaum)吸引子的无限自动机，在文献[44]中作出了旋转数为黄金分割数 0.618…的圆同胚映射的无限自动机。这些具体例子是我们工作的出发点。

对圆同胚映射我们知道这里最重要的量就是旋转数。当旋转数是有理数时，由映射生成的形式语言是正规语言；而当旋转数是无理数时，相应的语言是非正规语言(见文献[45])。由于在 0 和 1 之间的无理数有不可列个，因此在这些非正规语言中间就有不可列个语言的复杂性超出了乔姆斯基的层次体系，而达到了不可计算的程度。但我们所得到的无限自动机的结构是具有普适意义的，它对一切旋转数是无理数的圆同胚都同样适用。实际上，决定这个无限自动机结构的就是旋转数的连分数展开[25,44]。

对单峰区间映射的正规语言已有完整的讨论[2,46,47]。由此类映射决定的语言为正规的充分必要条件是映射的揉序列为终极周期序列。对非正规语言的讨论还只有零星结果。在文献[41,43]中讨论了费根鲍姆吸引子的形式语言，证明了它是上下文有关语言，但不是上下文无关语言。在

文献[48]中讨论了由同态生成的一大类非正规语言，其中包括由菲波那契(Fibonacci)序列加旋转生成的语言。最近利用在文献[49]中提出的揉映射和在文献[47]中的部分结果，我们找到了所有单峰映射生成的非正规语言的无限自动机的普适结构。

5. *概率自动机*

在自动机中引进概率因素，即自动机的状态转移不完全是确定性的，而是服从一定的概率，这是很自然的考虑。这方面的研究见文献[30,41,44]。但这里还有很多问题不清楚。理论上由于 BRS 测度(即所谓物理测度)可能不存在，因此概率自动机的理论基础还有困难。在文献[41,44]中提出了从测量到的符号序列重建自动机的方法，这有很大实际意义，但目前的计算量相当大，尚待改进。同时，这种方法在理论上也还需要作进一步的研究。

6. *元胞自动机*

检验各种复杂性度量的另一个系统就是元胞自动机。它的状态本身就是符号序列，这里不需要经过粗粒化处理。在一维元胞自动机和生命游戏中出现的突发行为，在一般的复杂系统中具有代表性[18,29]。关于这方面的文献可以参看论文集[50～52]。但是这方面的研究结果有很多是从计算机实验得到的，还缺乏严格的理论基础。用形式语言工具讨论元胞自动机时有很大的困难，严格建立的结果还不多，见文献[52～56]等。文献[57]代表了这方面的较新动向。

参考文献

[1] Li M, Vitányi P. An Introduction to Kolmogorov Complexity and its Applications. Berlin: Springer, 1993

[2] 谢惠民.复杂性与动力系统. 上海:上海科技教育出版社,1994

[3] Lempel A, Ziv J. *IEEE Trans IT*, 1976, **22**: 75

[4] Kaspar K, Schuster H G. *Physical Rev A*, 1987, **36**

[5] 郝柏林.从抛物线谈起.上海:上海科技教育出版社,1993

[6] Collet P, Eckmann J-P. Boston: Birkhäuser, 1980

[7] Devaney R L. An Introduction to Chaotic Dynamical Systems. 2nd Edition. Addison-Wesley, Reading MA, 1989

[8] May R M. *Nature*, 1976, **261**: 459

[9] de Melo W, van Strien S. One-Dimensional Dynamics. Berlin: Springer, 1993

[10] Mandelbrot B B. The Fractal Geometry of Nature. San Francisco: Freeman, 1982

[11] Gleick J. Chaos: Making a New Science. New York Viking, 1987
[12] Penrose R. The Emperor's New Mind, Concerning Computers, Minds and The Laws of Physics. Oxford: Oxford University Press, 1989
[13] Lorenz B N. *J Atmos Sci*, 1963, **20**: 130
[14] Sparrow C T. The Lorenz Equations: Bifurcations, Chaos, Strange Attractors. New York: Springer, 1982
[15] Berlekamp E R. Conway J H, Guy R K. Winning Ways, for Your Mathematical plays, Vol 2. New York: Academic Press, 1982. 817
[16] Ueda Y. *Nonlinear Science Today*, 1992, **2**(2): 1
[17] Peliti L, Vulpiani A. Measures of Complexity. New York: Springer, 1987
[18] Roetzheim W. Enter the Complexity Lab. Sams Publishing, Indianapolis IN, 1994
[19] Wackerbauer R, Witt A, Atmanspacher H, Kurths J, Scheingraber H. *Chaos, Solitons & Fractals*, 1994, **4**: 133
[20] Herzel H. *Syst Anal Model Simul*, 1988, **5**: 435
[21] Xu J H, Liu Z R. *Chaos, Solitons & Fractals*, 1994, **4**: 2111
[22] Weisbuch G. Complex Systems Dynamics, An Introduction to Automata Networks. New York: Addison-Wesley, 1991
[23] Hopcroft J E, Ullman J D. Introduction to Automata Theory, Languages, and Computation. Addison-Wesley, Reading MA, 1979
[24] Davis M D, Sigal R, Weyuker E J. Compatability, Complexity and Languages, 2nd Ed. Boston: Academic Press, 1994
[25] Hao B L. Elementary Symbolic Dynamics and Chaos in Dissipative Systems. Singapore: World Scientific, 1989
[26] 郑伟谋,郝柏林.实用符号动力学.上海:上海科技教育出版社,1994
[27] Brudno A A. *Trans Moscow Math Soc*, 1983, **2**: 127
[28] Keller G. Lecture Notes in Math. 1486. Berlin: Springer, 1991: 216
[29] Wolfram S. *Commun Math Phys*, 1984, **96**: 15
[30] Grassberger P. *Int J Theor Phys*, 1986, **25**: 907
[31] Grassberger P. *Z Naturforsch*, 1988, **43**a: 671
[32] Grassberger P. Peliti L, Vulpiani A (eds). Measures of Complexity, Lecture Notes in Physics 314. Berlin: Springer, 1988. 1
[33] Hao B L. *Physica D*, 1991, **51**: 161
[34] Günther R, Schapiro B, Wagner P. *Chaos, Solitons & Fractals*, 1994, **4**: 635
[35] 谢惠民.动力系统的语言复杂性.非线性动力学学报,1995,**2**:223
[36] Morse M, Hedlund G A. *American J of Math*, 1938, **60**: 815
[37] de Luca A, Varricchio S. Capocelli R M (ed). Sequence. Berlin: Springer, 1990, 258
[38] Xie H M. *Complex Systems*, 1995, **9**: 73
[39] Li T Y, Yorke J A. *Am Math Monthly*, 1975, **82**: 985
[40] Rozenberg G, Salomaa A. The Mathematical Theory of L Systems. New York: Academic Press, 1980
[41] Crutchfield J P, Young K. Zurek W H(ed). Complexity, Entropy and the Physics of Information. SFI Studies in the Sciences of Complexity, Vol. VIII. Addison-Wesley,

Reading MA 1990: 223
[42] 陈曦,卢钦和,谢惠民. 数学进展, 1992,**22**: 185
[43] Moore C. *Nonlinearity*, 1991, **4**: 199
[44] Crutchfield J P. *Physica D*, 1994, **75**: 11
[45] Friedmann E J. *Complex Systems*. 1991, **5**: 335
[46] Wang Y, Xie H M. *Nonlinearity*, 1994, **7**: 1419
[47] Xie H M. *Nonlinearity*, 1993, **6**: 997
[48] Xie H M. *Nonlinearity*, 1996, **9**: 1469
[49] Hofbauer F. *Mh Math*, 1980, **90**: 117
[50] Wolfram S. Theory and Applications of Cellular Automata. Singapore: World Scientific. 1986
[51] Gutowitz H(ed). *Physica D*, 1990, **45**
[52] Jackson E A. Perspectives of Nonlinear Dynamics. Cambridge: Cambridge University Press, 1990, 2
[53] Hurd L P. *Complex Systems*, 1990, **4**: 119
[54] Hurd L P. *Complex Systems*, 1990, **4**: 131
[55] Hurd L P, Kari J, Culik II K. *Ergodic Theory and Dynamical Systems*, 1992, **12**: 255
[56] Kari J. *SIAM J Comp*, 1992, **21**: 571
[57] Hanson J E, Crutchfield J P. *J Stat Phys*, 1992, **66**: 1415

(谢惠民)

第4章　生　物　膜

§4.1　膜　蛋　白

膜蛋白指与生物膜相关联的蛋白质，虽然它也和溶液中的蛋白质同样都由不同的氨基酸缩合而成，有它的一级、二级、三级以至四级结构。然而，至今它的研究进展却远远落后于溶液中的蛋白质。从原子分辨水平上的三级结构来比较，至1995年，溶液中蛋白质的空间结构研究清楚、并已储存到国际数据库的蛋白质数超过3 100种，而且几乎以每天增加一个结构的速度迅速发展。与之形成对照的是，膜蛋白则从1985年第一次完整解析了紫菌光合反应中心以来，10年内在原子分辨水平上结构研究清楚者不超过10个。造成这种差距的原因不外乎两个方面：一是膜蛋白所处的环境复杂，除其中一部分处于膜两侧水相以外，相当一部分处在膜脂双层疏水环境之中，对这种蛋白质既要获得适于用三级结构方法研究的对象（主要是晶体），又要保持其原有构象十分困难，人们甚至还没有找到制备成适当晶体方法的规律。二是膜蛋白本身组成复杂，分子大，并常常多次跨膜，或者由多个亚基组成，这就大大限制了现有的几种空间高分辨结构研究方法的应用。

然而膜蛋白结构的研究却是从微观（分子）水平了解生命基本过程的关键。无论是细胞内外物质的通透，信息的跨膜传输，还是细胞内能量的转换无一不在膜内进行。在完成功能过程中膜蛋白一般都起着主要角色的作用，例如通道、受体、电子传递链等等。今后对生命基本过程的了解，必须搞清楚膜蛋白结构及其转变。因此可以不夸张地说，未来结构生物学中将以更大的精力注重膜蛋白的研究，因为它是从单纯溶液状态向细胞活动过渡的极其重要的一环[1~4]。

4.1.1　高分辨膜蛋白结构研究

现在一提到膜蛋白结构似乎就是指原子水平的空间结构，当然在结

构生物学中这仍然是最主要的部分。然而借鉴溶液中蛋白质常用的几种研究方法，应用于膜蛋白时既有成功的一面，也有欠缺之处。

1. X射线衍射晶体分析

对于膜蛋白，重要的是制备成适于X射线衍射用的稳定晶体。膜蛋白基本处于疏水的膜脂环境之中，脂和蛋白质共结晶是一个困难问题。目前常用的方法是以适当的去污剂取代脂，这种取代需要一定的技巧。国际上第一次结晶成功并获得晶体分析的是紫菌(*Rhodopseudomonas viridis*)的光合反应中心，相对分子质量为150 000，分辨率0.3 nm。这项工作在膜蛋白研究中成为一个里程碑，其完成者J. Deisenhofer，R. Huber与H. Michael于1985年，即结晶成功后3年，被授予诺贝尔化学奖。这项研究分析出光合反应中心分为细胞色素、H、M、L等四个亚基，其中L与M位于膜脂区中，各有5个α螺旋，H则只有一个α螺旋，辅基与各亚基间的相互关系也已弄清楚。因而对光合作用中的电子传递过程有了较深入的了解。

此后，在膜蛋白晶体衍射方面逐步取得了一些进展，见表4.1.1。

表4.1.1 用X射线衍射法研究过的膜蛋白

时 间	主要完成者	样 品	相对分子质量	分辨率(nm)
1985年	Deisenhofer	紫菌光合反应中心 (*R. viridis*)	150 000	0.3
1990年	Henderson	嗜盐菌菌紫质 (bacteriorhodopsin)	26 000	0.35
1991年	Weiss	通道蛋白 porins (*Rhodobactor capsulata*)	124 000	0.18
1994年	Picot	前列腺素H合成酶 (prostaglandin H synthetase)	~70 000	0.35
1995年	Yoshikawa, Michael	细胞色素C氧化酶 (cytochrome C oxidase)	200 000	0.28

我们不可能逐个加以叙述，但有几点值得提到的是：①膜蛋白晶体衍射研究的速度随着经验的积累将逐步加快，特别是1995年8月同时在*Nature*和*Science*杂志上刊登的、由德国和日本两个研究小组都得到了细胞色素C氧化酶(cyt. C oxidase) 0.28 nm的研究结果，成为近年膜蛋白研究的重要进展[5,6]。②表4.1.1中五类蛋白中在膜内部分大多数都以

α 螺旋为主，但通道蛋白膜孔蛋白(porins)则以 β 折叠为主，几个 β 片层彼此以氢键连结，并首尾相连形成 β 筒状(β-barrel)。这是 α 螺旋以外的第二种形式，并常以三聚体形式存在。前列腺素 H 合成酶采取第三种形式，即不跨膜而以部分弯折进入膜内(monotopic)。此外，是否还可能有其他方式，尚待进一步研究。③每一类并不止一种，例如 porins 根据基因来源不同有 OmpF、OmpA、PhoE 等几种，但总数约在 10 个左右。

X 射线衍射法的缺点是需要尺寸较大(毫米数量级)的晶体，并在较长时间内稳定，所得结果是一个时间平均的稳态结构等等。用同步辐射及改进探测手段，结合其他技术的研究结果将是可能的发展方向。

2. 电子晶体学——二维晶体的电镜三维重组法

三维晶体的困难部分可用先组成二维晶体，然后用三维重组的方法克服。因为二维晶体获得的成功率较大，要求晶体的尺寸也远小于三维晶体。事实上，上述嗜盐菌菌紫质早在 1975 年就已用这一方法获得 7 股螺旋的模型，分辨率在当时已达到 0.7 nm，而且嗜盐菌菌紫质的二维晶体在天然条件下即可获得。在三维重组时，需在电子显微镜中改变样品平面的倾角，以获取多个衍射面的数据；对于有规则样品，则一个平面衍射数据获得后即可根据样品的规律性推断三维结构。虽然目前此法的分辨率还在 1～2 nm 间(个别样品如菌紫质较高)，但随着技术的改进(包括冰冻电镜的应用)和晶体样品的完善化，将有希望成为膜蛋白结构研究的主要手段。目前用此法研究成功的例子已有不少，如捕光色素 LHC-I 与 II，乙酰胆碱受体，神经细胞膜配体门控的离子通道、细胞色素还原酶、Ca^{2+}-ATP 酶、Na^{+}-K^{+}-ATP 酶。近年来获得成功的还有红细胞带 III 蛋白和水通道(aquaporin)(1993)，如红细胞膜上的 CHIP28(channel-forming integral protein，相对分子质量 28 000)是对水特异的通道。

3. 多维核磁共振

核磁共振，特别是二维核磁共振(2D-NMR)已成为独立于 X 射线晶体衍射之外的新的高分辨测定结构的手段。但它主要用于溶液中的蛋白。对于膜蛋白而言，由于处于脂环境，脂分子上的氢核干扰太大，很难获得结果，更何况目前能测定结构的溶液中蛋白质相对分子质量也不超过 20 000，加上其他核标记及方法学上改进，也不会超过 30 000。而用氘取代全部脂又极为昂贵。但对于表在蛋白(peripheral protein)，特别是一些

与膜作用的多肽则是适用的手段。已研究过的有短杆菌肽A和S、丙甲霉素(alamethicin)、蜂毒素(melittin),以及一些小肽如激素、神经肽(内啡肽、强啡肽、脑啡肽)、P物质等。

由此可见,在技术上,目前在高分辨结构研究中仍然期望新技术的开发与应用。近年来扫描探针显微术(scanning probe microscopy)的应用已显露出一定的优越性。我们在近年来不仅用扫描隧道显微术(STM)研究了DNA,而且研究了脂的多型性,并获得成功。更重要的是,对膜蛋白似乎也有一定价值,例如观察到插入脂质体的免疫球蛋白,细胞色素C氧化酶,抗原与抗体的作用等等。国外目前原子力显微镜(AFM)的应用日益增加。但无论STM还是AFM,其分辨率还有待提高,特别是在应用于膜蛋白研究之时。

4.1.2 其他研究途径的重要性

在高分辨结构研究的同时,我们也不能忽视其他各级结构的探讨。这并不完全因为三级结构的研究困难太大,退而寻求其他途径,而是因为结构生物学研究的终极目的在于阐明结构与功能的关系。完成某种特定功能,有时需要涉及分子的整体,有时则只需了解其局部变化即可。换句话说,分子中个别部分在完成功能中的重要性并不一定相同。例如酶的作用机制中,活性中心的变化显然大于其余部分。功能作用完全不同的超氧化物歧化酶(SOD)和免疫球蛋白的抗原结合部位(F_{ab}的可变区)的基本结构是完全相同的,可见在产生歧化反应和与抗原结合时起主要作用的只是分子中的一小部分。即使像O_2进出肌红蛋白内部的血红素基团,需要分子整体开合,但成为能障的主要是由于苏氨酸E_{10},组氨酸E_7和缬氨酸E_{11}。因此,应用各种方法探索关键部位的构象变化就成为膜蛋白研究的重要部分。其中包括各种光谱方法,如吸收光谱、荧光光谱、圆二色谱(CD)、红外光谱、特别是傅里叶变换红外(FTIR)、拉曼光谱等;波谱技术如ESR,NMR等;分子生物学方法,如定位突变法,cDNA克隆法;辐射失活法;拓扑结构的一些分析方法,如基因融合、化学修饰法、抗体结合法等等。

用提取的膜蛋白分析其一级结构、或者用cDNA克隆了解膜蛋白的一级结构,到膜蛋白三级结构的认识,实质上也是一个折叠的问题,也就是从序列到结构的研究。一个比较常用的手段是分两步走,先从序列到二

级结构，再从二级结构间的相互作用了解空间结构。由此可见对二级结构的研究在膜研究中占有重要地位。近十几年来，FT - IR 或 FT - Raman 谱的研究起到重要作用。由于分辨率的提高，去卷积(deconvolution)、二阶导数谱(second derivative)以及拟合方法的计算机快速应用，使原来包含在一个由各种二级结构混合形成的酰胺峰(amide I,II 等)能明显分开，从而解析其中的各种成分及其含量，使膜蛋白中各种二级结构的研究及其在功能过程中的改变有了较有效的手段。

1994 年，W. L. Hubbell 与 C. Altenbach 所发展起来的所谓定位自旋标记法(site - directed spin label，简称为 SDSL)成为测定膜蛋白拓扑结构的有效方法。它把含 NO 基的氨基酸取代天然蛋白的特定部位，用一对自由扩散的顺磁探剂的 NO 侧链分别放在亲水部位与疏水部位，测定其碰撞频率。极性探剂在亲水部位或非极性探剂在疏水部位时，碰撞频率都高；而在蛋白内部者则低。如果用不同部位标记物进行 NO 扫描则可见到碰撞频率的周期性变化，对 α 螺旋，周期为 0.36 nm，而对 β 结构则为 0.2 nm，正好反映了 α 与 β 的二级结构。它在膜蛋白中的应用打开了直接测定的新途径。

此外，用荧光标记法、荧光猝灭法测定膜蛋白中二级结构的变化，用定位突变法测定个别残基的改变对插膜与功能的影响等等都在膜蛋白结构与功能关系研究中起到了重要作用。

综上所述，当前对膜蛋白的研究还不能依靠单一技术，而要各种水平(从一级到四级结构)和各学科(物理的、化学的和生物学的)方法的综合；需要静态研究方法与动态的(时间分辨的)方法相互结合；此外还需要大量实验研究和理论分析方法的互相促进。

4.1.3 膜蛋白构象研究中的一些理论问题

如前所述，膜蛋白构象的研究也是蛋白质结构从序列到空间结构这一结构生物学普遍问题中的一个组成部分，亦即结构预测的部分。和一般溶液中蛋白质折叠问题既有相似的一面，又有不同之处。虽然膜蛋白一般来说比较复杂，但从已经了解的一些结构看，膜内的疏水环境对膜蛋白构象产生一定的局限作用，其中 α 螺旋、β 筒等似成为可以允许的少数几种结构，因而复杂性大为降低。再从整个膜蛋白与膜的拓扑关系观察，一级结构的 N 端与 C 端可能位于膜脂双层内外两侧，而跨膜部分的疏水性一

般较强，而且跨膜段之间亲水性较强的部分能以环(loop)的形式分布在膜的内外两侧。由此可见，分析整个膜蛋白的一级序列，找出其中疏水性较强的片段(当然还需要考虑跨膜段形成二级结构所需的最少残基数)，是从理论上分析可能跨膜方式的第一步。比较困难的是第二步，即 α 螺旋之间的相互作用如何形成螺旋束。这种相互作用的动力来自氢键形成，还是电荷相互作用，存在着相当多的不确定性。膜内螺旋常见脯氨酸存在，而在该处螺旋常呈扭曲，它对螺旋相互作用的贡献也是当前研究的一个重要方面。总之，结构预测的第二步，目前远未解决。

膜蛋白构象的分子动力学研究显然也是理论物理学家可以发挥作用的一个方面。关于分子动力学计算在蛋白质结构研究中的应用，已在其他报告中有较详细的介绍。这里仅指出一点，即对于溶液中的蛋白质，典型折叠时间在秒甚至分的数量级，因此直接模拟包含了许多纳秒级的过程，两者相比差距太大。而对于膜蛋白，这一问题有可能克服，即只指明二级结构，再讨论相互作用的两步法。对于后者用分子动力学方法已获得成功，特别是应用于嗜盐菌菌紫质。O. Edholm 和 F. Jähnig 应用标准场方法，但对脂与水溶剂的差别增加了一个疏水势，同时增加一项氢键项以增强氢键作用，获得了较好的结果。需要指出的是，任何模型化的工作都要和实验相比较，才能获得有成效的结果。

膜蛋白结构研究的一般问题在前面已作了一般性介绍，但对于特定的膜蛋白，今后的工作还应结合该膜蛋白的特殊功能深入探讨。例如对于通道，应着重研究通道的选择性与控制通道启闭的分子机制与通道构象变化的关系；对于受体，应着重研究配体-受体作用的特异性以及结合时构象互相适应改变与信息传递的关系；对于电子传递链，应着重研究电子传递过程中的构象变化与能量转移关系等等。从目前来看，研究上述问题首先要解决的是膜蛋白的结构本身。由于学科发展迟缓，所得到的结果数量还太少，因此今后一段时期内膜蛋白结构的研究仍将成为重点，以便为这一领域奠定更坚实的基础。

参 考 文 献

[1] McCarthy M P, Yeagle P ed. *The Structure of Biological Membranes*. CRC Press, Boca Raton, Ann Arbor, 1992. **395**

[2] Kühlbrandt W, Gennis R B. Membrane proteins, editorial overview, *Current opinion in structural biology*, 1994, **4**: 517

[3] Traxler B, *et al*. The topological analysis of integral cytoplasmic membrane proteins. *J. Memb Biol*, 1993, **132**: 1

[4] Edholm O, Jähnig F. Molecular dynamics simulations of bacteriorhodopsin. In: Pullman A, *et al*. eds. *Membrane proteins: structure, interactions and models*. Dordrecht: Kluwer Academic Publishers, 1992, **47**

[5] Iwata S, *et al*. Structure at 2.8Å resolution of cytochrome C oxidase from paracoccus denitrificans. *Nature*, 1995, **376**: 660

[6] Tsukihara T, *et al*. Structures of metal sites of oxized bovine heart cytochrome C oxidase at 2.8Å. *Science*, 1995, **269**: 1069

(林克椿)

§4.2 生物膜液晶模型

4.2.1 生物结构与液晶

1888年,奥地利植物学家F. Reinitzer在研究胆甾醇酯对植物的作用时,发现这种有机材料竟然有两个熔点:在145.5℃,胆甾醇酯先熔化成一种混浊的液体;到178.5℃,液体突然变成全透明。"如果说固体的胆甾醇酯已在145.5℃熔化为液体,那么,又是什么不同于固体与液体的鬼东西在178.5℃熔化?"为了弄清这个问题,他把这种神奇的材料寄给德国卡斯鲁尔大学的物理教授O. Lehman(用附有热台的偏光显微镜研究晶体的鼻祖)。在对这种材料进行充分研究后,Lehman认识到,混浊态的胆甾醇酯是一种"几乎是流体的软晶体",或者更精确地说是一种"晶体结构的流体"。正是在生物学家与物理学家的合作下,物质存在的一种新态——液晶于19世纪的末期被人类发现了。

其实,生物结构与液晶的联系可追溯到19世纪中叶。对细胞学说创立有重要贡献的Virchow1854年就发现神经细胞髓磷脂溶液具有偏光性(即液晶的特性)。Lehman在发现液晶后不久,即专文[1]论述液晶在生命科学中的重要前景。许多生物学家在液晶研究早期都对此发表过评论。在1933年的法拉第液晶讨论会上,生物结构的液晶性质便正式被提出来讨论,而研究液晶的科学家本身就是著名的生物学家也不鲜见。如发现近晶相(smectic phase)液晶焦锥结构的F. Grandjean[2,3]就是一个著名的发育生物学家,写过有关生物发育机制方面的重要论文[4,5]。胚胎学家李约瑟就曾把哺乳动物胚胎发育过程中肢节轴索的诞生与液晶相变产生的几何拓扑结构联系在一起。50年代,生物学家便已逐渐认识到肌肉组织

与细胞结构都显示出与液晶各种相(从向列相 nematic,胆甾相 cholesteric,到近晶相)相似的分子堆积结构,并出版过有关专著[6]。1965 年在美国肯特州立大学召开的第一届国际液晶会议上,生物结构的液晶性质便正式成为会议的一个独立的议题。到 1974 年在斯德哥尔摩召开的第五届国际液晶会议上,人们特别把生物膜与溶致液晶(lyotropic)挂起钩,会后 Friberg 写出《溶致液晶与生物膜》一书。稍后不久,液晶国际会议的发起人 Brown 和 Wolken 便出版了广为流传的单行本《液晶和生物结构》。

李约瑟曾有一句名言:生命现象可从液晶的分子结构想象到。这里举一个例子说明生物学中被认为是难题的一些现象,可从液晶性质中得到解答。大家都知道迈克尔逊是著名的光速测量的研究者,但少有人知道他在研究金龟子甲壳光学性质方面也曾有过贡献。在 1911 年,他通过测量,证明这种昆虫甲壳反射的太阳光是圆偏振的[7]。昆虫甲壳能够从不同角度反射出不同颜色的光,这曾经使生物学家长期疑惑不解[8]。昆虫甲壳反射光圆偏振性及颜色与反射角的依赖性的探索成了一个难题。直到 1951 年,de Vries 在研究胆甾相液晶的光学性质时,这个问题才得到彻底解答[9]。下面要详细介绍的生物膜液晶模型也是从生物膜研究遇到的一个难题开始的:为什么人体中的红细胞是双凹碟形。这个由理论物理学家提出的模型不仅解开了血细胞形状之谜,而且从理论上预见了迅速为实验证实的生物膜新构形—环状膜泡;还对生物膜的螺旋结构——在生命体系中最常见并最难理解的膜构形,提出了令生物化学家、医学家信服的理论描述。

4.2.2 液晶解开红细胞形状之谜

为什么人体中的红细胞是双凹碟形而不是其他几何形状?这个问题与显微镜的发现一样古老。哺乳类动物与人的红细胞是无核细胞,其形状完全取决于细胞膜的力平衡,是研究生物膜的理想客体,因此人们对红细胞形状特别感兴趣。为解释红细胞的双凹碟形,生物力学家曾从薄壳力学的概念出发,给出各种计算模型。如冯元桢与 Tong(1968)曾把红细胞想象成由弹性薄壳夹一层各向同性液体的“三明治”,为达到期望的双凹形状,他们假设膜厚度是不均匀的,但这有悖于电镜的观察结果——膜厚度实际是均匀的。尤其是与红细胞沿血管壁的坦克履带式运动方式比较,薄壳模型是不成立的。为此,Sirs(1970)提出生物膜是由一片片簧片式元件

连成的"甲衣"，犹如手表的簧带，但其计算结果与实际形状不合(Branto，1969；Weinstein 和 McNutt，1970)。Lopez 等人(1968)把双凹形状归结为膜带电不均匀，但 Greer 和 Baker(1970)的测量表明膜带电是均匀的。Shrivastav 和 Burton 则假设红细胞膜相向部分存在长程吸引力，红细胞轴心附近膜的内凹则是由于这种力相吸的结果，但电镜研究否定这种力的存在。Murphy(1965)把双凹的成因归于胆固醇的分布不均匀，但实验(Seeman 等，1973)发现胆固醇在膜上的分布是均匀的。有鉴于这些欠妥的模型，加拿大生物力学家 Canham(1970)抽象地假设细胞的形状只取决于弯曲弹性能：$E=\int(2H)^2\mathrm{d}A$。这里 H 是膜曲面的平均曲率，积分遍取膜的表面。Canham 利用计算机模拟，果真发现双凹碟形是弹性能的极小值状态。不幸的是，两位液晶物理学家 Deuling 与 Helfrich 1976 年进一步研究，发现哑铃状的细胞也是这个模型的结果。但人们并没有发现有哑铃状的红细胞。截止 70 年代，红细胞的双凹碟形问题仍未解决，而成为生物力学的一个难题。

生物膜的主要结构分子是类脂化合物和磷脂，它们都是双亲分子。按照公认的流体镶嵌模型(Singer 和 Nicolson，1972)，这些脂类分子以双分子层成膜，其亲水极性基团朝外与水接触，而两条疏水烃链夹于膜内。在正常生理温度，它们可以在膜上自由流动。这种排列方向有序的流体正是分子处于液晶态的宏观特征。因此，1973 年液晶理论物理学家 W. Helfrich 把类脂分子的烃链取向看作单轴液晶的指向矢，那么生物膜即可描述为"盒厚"约为两倍脂类分子长度的垂面校列向列相液晶盒，其厚度约在 8 nm 数量级。把膜面法向(即在正常生理温度 L_α 相类脂分子烃链取向)取为液晶指向矢，借助液晶弹性自由能 Frank 表达式[10]，Helfrich 导出单位面积细胞膜的弹性能是

$$F=(k/2)\int(2H+C_0)^2\mathrm{d}A+\bar{k}\int K\mathrm{d}A。\qquad(4.2.1)$$

与 Canham 上述自由能比较，(4.2.1)式多了一项高斯曲率 K 的积分以及与平均曲率 H 线性积分相关的自曲率常数 C_0。后者是由液晶的自展曲 S_0 导出的[11]。它可以描述膜两侧环境，如膜电势与溶液的 pH 值，也可以是描述膜内外两层分子层成分的不对称性的综合参数。如胆固醇含量，所镶嵌蛋白质、糖脂分布。对于闭合膜泡，生物膜形状自由能还应加上溶

液在膜内外形成的渗透压差 $\Delta p=p_0-p_i$ 引起的体积能 $\Delta p\int dV$ 与膜表面张力的表面能 $\lambda\int dA$，λ 是张胁强。利用数值计算，Deuling 与 Helfrich 在 1976 年证明在 C_0 为一定负值时，红细胞双凹碟形是膜自由能的唯一极小值态[12]，从而使这一生理难题在液晶模型中得到解决。同年 Helfrich 获得欧洲物理学会的惠普奖。然而，这些为物理学界认可的数值计算结果并不能一下子为生物学家所接受。在 1987 年举行的中日美国际生物力学会议上，人们仍然为生物膜是纯流体抑或固体薄壳这两种对立的观点争论不休。由 Hotani 领衔的日本科技振兴学会一个旨在研究生物膜构形与分子机械的小组的成员，在 80 年代研究红细胞形状仍然沿用无 C_0 的 Canham 模型。为使生物膜液晶模型为科学界广泛承认，有许多艰苦的研究仍需进行。这就为我国理论物理工作者提供了机遇。

首先，1987 年笔者与 Helfrich 对 Helfrich 的自由能进行变分，而导出了生物膜形状的普遍方程[13]：

$$\Delta p - 2\lambda H + k(2H + C_0)(2H^2 - 2K - C_0H) + 2k\nabla^2 H = 0。\tag{4.2.2}$$

从这个方程出发，笔者与日本合作者[14]于 1993 年在国际上首次给出红细胞双凹碟形的解析解

$$\begin{cases} z(\rho) = z(0) + \int_0^{\rho}\tan\psi(\rho')\mathrm{d}\rho', \\ \psi(\rho) = \arcsin[C_0\rho\ln(\rho/\rho B)], \end{cases}\tag{4.2.3}$$

其中的 $z(\rho)$ 是细胞沿对称轴剖面的轮廓线，ρ 是距红细胞膜垂直旋轴的距离，$\psi(\rho)$ 是膜面的切角。用初等微积分方法即可证明，在 C_0 为负值时，曲面(4.2.3)正是双凹碟形。(4.2.3)式解有一个限制，$\Delta p=\lambda=0$。这个条件正是生物化学家 Tanford[14]对磷脂膜所提出的物理化学条件。

4.2.3 环形生物膜泡

欲使生物学界对液晶膜模型感兴趣，人们需要更多有说服力的事例，或者说应有惊人之举。1990 年笔者从理论上预见方程(4.2.2)有一个环形曲面解[15]，该曲面可以看成一个小圆(半径为 r)绕一个大圆(半径为 R)转出来的曲面。一个苛刻条件是半径比 $R/r=\sqrt{2}$ 或无穷大。次年，这个预言即被巴黎高师统计物理实验室的 Mutz 和 Bensimon 全面证实[16]，

他们在文章摘要中即宣布他们的实验证实了文献[15]的预言，该实验室观察到 24 个大小不等的部分聚合的磷脂环形泡。其严格遵守 $\sqrt{2}$ 半径比率，并且这 24 个泡都满足文献[15]所给出的自曲率的稳定条件 $C_0r_0<-3.9$。在这里 r_0 是与泡等面积的球面之半径。

关于 $R/r=\infty$ 的管状环形泡的预言也很快被美国海军实验室生物分子科学与工程中心的 Rudolph 等人发表在 *Nature* 上的有关磷脂膜的实验观察所证实[17]，在他们文章摘要中引用的两篇理论预言文章也是笔者所撰[13,15]。

1992 年，美国与以色列联合研究组的 Z. Lin 等人也在活性剂分子的胶囊体系中发现了环形泡[18]，这在胶囊化学中也是罕见的。诚如他们文章引言中所说“过去绝大多数的工作都集中在球拓扑(所谓的零亏格)泡，而理论建议存在着环形和其他亏格的泡[3~5]”。这里他们引用的三篇理论文献中，文献[15]是开篇之作。

环形泡的理论发现与实验证实的重要意义是直接证明了生物膜本性为液晶这一基本观点。早在本世纪初期，Friedel 与 Grandjean[2,3]就发现近晶相液晶会形成一套套环面相套叠的焦锥结构(focal conic texture)，它们是几何上的杜邦环面系。显然，生物膜的环形泡与近晶相的焦锥结构的本质及机制是一致的——液晶是其共同属性。想了解生物膜上发生的事，最好是先想一想液晶会发生什么事。这一点远远超越了 Helfrich 的流体膜理论。下文将给出一个更生动的例子。

4.2.4 胆甾相液晶与生物膜的螺旋结构

在千姿百态的生物形态中，小若细菌大如海螺，螺旋形状非常多见。令人迷惑不解的是从生物分子上溯到高等动物，各个层次都存在着普遍的手征不对称。例如，地球上发现的氨基酸多见于左旋，蛋白质和 DNA 基本上是右旋。螺形细菌、旋花类植物的螺旋盘绕以及海螺的外壳也均以右旋占多数。理解不同层次生物结构对某种旋向的偏好以及探讨不同层次手征性的因果关系，是探索生命起源与进化奥秘最有趣的问题之一。最近手征分子构成生物膜的螺旋结构的发现与理论研究对上述问题作出了部分回答。

天然生物膜的螺旋结构已早有报道(如我国学者林克椿 80 年代初期在国外的工作)。合成人工生物膜的螺旋结构从 80 年代中期进入研究高

潮。1984年日本化学快报同时报道了两个实验室对同一类手征分子组成的生物膜的螺旋结构[18,19]。在较高温度,这类膜以闭合泡形式出现。降低到某一临界温度以下,泡解体,组成膜的分子重新凝聚为一种扭曲螺带。随着温度的进一步降低,这些螺带会变成如同缠绕在一个长圆柱上的缎带。这螺带不仅沿轴向迅速生长,而且也会沿侧向生长使螺间距闭合而长成半径为0.1～10 μm的长管子。1990年,美国海军实验室的研究人员发明了在这种生物膜管上镀金或铜的新技术。这种用生物工程进行微结构加工的技术可用于光电子技术、微形手术和药物导弹,显示了这种结构潜在的应用前景。因此,弄清这种结构的机理十分迫切。

有三种典型的螺旋膜理论:第一种是1991年诺贝尔物理学奖得主de Gennes关于膜自发电极化的卷曲模型[20];二是Lubensky和Prost关于非手征性柱形膜泡由于弯曲弹性能与膜边缘线张力竞争的生长模型[21];三是Helfrich和Prost关于分子倾斜膜的提法[22],并由我们从胆甾相液晶弹性能理论予以精确化的理论[23]。我们首先注意到,Helfrich流体膜理论(等式(4.2.1))只是建立了非手征性的向列相液晶的弹性理论,所以导出的方程(4.2.2)是在左右手坐标变换中不变的,因此不可能有螺旋结构的解。而液晶分子若有手征性,在降低温度时,在某个相变温度时,会演变成胆甾相,该相有三维的螺旋指向矢分布。根据这个原理,我们从胆甾相液晶自由能导出了手征分子倾斜膜的弹性自由能。对其变分,我们得到了描述这种膜的新的形状方程,以及描述膜分子倾斜方向在膜面上的投影——膜上的线场方程。结果发现,实验上观察到的扭曲螺带与盘旋形螺带正是所得方程的精确解。尤其我们证明了螺管半径与胆甾相液晶的螺距是同一数量级(0.1～10 μm),再次由生物膜的螺旋结构找到其与液晶的对应。

对上述三种理论,美国海军实验室的生化专家在*Science*[24]与*Phys. Rev Lett*[25]上评论道:首先,添加电解质对溶液中的螺管半径不起作用的实验(Chappell与Yager,*Biophys. J.* 60,1991,952)与de Gennes的理论不合,因为照理膜卷曲的静电引力应受到电解质的屏蔽。第二,螺管半径r与长度L无关联的实验(Geoger等人,*J. Am. Chem. Soc.* 109,1987,6169),同时排除了de Gennes理论与Lubensky－Prost理论,因为前者预言了$r \sim L^{1/3}$,而后者预言$r \sim L^{1/2}$。这样只有第三种理论可以与实验相

符。

对于在临床医学工作领域的生物学家，他们又是怎样来看待这些由理论物理学家用复杂数学工具(如微分几何)建立起来的手征分子倾斜膜理论？由美国科学院院士 G. Benedek 领导的哈佛医学院与麻省理工学院物理系组成的一个联合研究组，对胆结石病理研究中发现的胆固醇螺旋结构进行了实验与理论探索。他们 1993 年发表在《美国科学院院报》的论文[25]在螺管形成的理论工作只引用了三篇文献，就是 Helfrich(1986)，Helfrich-Prost[22]，和我们的论文[23]。这说明，在物理学家与生物学家联手探讨生命奥秘的战线上，理论物理的分析性结果已开始得到重视。

4.2.5 讨论

本世纪特别是本世纪下半叶，生命科学发展最明显的特点是生物学正改变着自己以描述性为主的面貌，逐步发展成为一门实验性和分析性的学科。在生物学的这一重大历史转化中，物理学、化学、数学以及技术科学向生物学的渗透起了关键性的作用。物理学、化学与生物学的结合，将推动生命科学更加激动人心的时代的到来。即将到来的 21 世纪，已被许多有远见的科学家预言为“生物学的世纪”。彭桓武先生在“对 21 世纪物理学的发展的一点猜想”一文(《物理》，1993 年第 3 期)中认为，物理科学与生物科学相结合是下一世纪物理学发展的两大趋势之一。这个预言实际上已显现了越来越多的征兆，有关液晶生物膜等复杂流体文章已从《物理评论 A》分册中分化出来，而从 1993 年开始诞生了《物理评论 E》的新分册，这个新分册到 1994 年已从原来每月一分册演变为每月两分册，有赶上专登载固态物理文章的《物理评论 B》的势头。著名物理学家 de Gennes 实际上也正在向这个趋势转化，他最近的新著《高分子动力学导论》是一本仅有 57 页的小册子，在书中四章的论述中，有一半就是专门探讨多肽(生物大分子)折叠与药物效应的。著名的晶体形状研究专家 M. Wortis 从 90 年代起也转行研究液晶生物膜理论，并在 1995 年起出任《物理评论快报》的生物物理专栏的主编。研究液晶生物膜模型几年来，我们都得到彭桓武先生的鼓励与亲临讨论指导。他特别嘱咐我们一定要跟踪 de Gennes 的研究手法，以简洁的理论方法去解决复杂的新问题。彭先生在《生物凝聚态》一文中，以更高的纲领性的概括与总结，为我们将来对理论生物物理学的研究指明了方向：生物学研究结构与功能，物理学研究

结构与性质，在这两者之中应该存在着反映功能的性质。因此，结构—性质—功能将是生物凝聚态研究的线索与模式。

从我们近年来研究液晶生物膜模型的进展及国际上的反响来看，生命世界的现象尽管纷繁复杂，但经过一定的"粗粒化"，其规律是可以简化到用连续介质模型理论来描述的。在结构生物学的研究热点中，大量的工作正集中在从分子水平、基因密码序列的一级结构向高级结构的探索，似乎不这样做就不是生命科学的主流。但从我们对螺旋膜结构的液晶模型研究来看，生物大分子的 α 螺旋与 β 折叠的奥秘仍然存在着可用"一维液晶模型"加以探索的途径。上述 de Gennes 的新著实际上已隐含着这种倾向。总之，一方面我们要从分子水平、基因角度来了解生物体，另一方面又要从生物凝聚态的角度来了解生物体。这应成为当前生命科学所面临的两个方面的重大课题。

参 考 文 献

[1] Lehman O. Flussige Kristalle und die Theorien des Lebens. *J Ambr Barth*, Leipzig, 1908

[2] Friedel G, Grandjean F. *Bull Soc Fr Mineral*, 1910, **33**: 192, 409

[3] Grandjean F, Hebd C R. *Seances Acad Sci*, 1917, **164**: 280, 431, 636

[4] Grandjean F. *Arch Sci*, 1957, **10**: 477

[5] Grandjean F. *Acarologia*, 1972, **14**: 454

[6] Frey - Wyssling A. Submicroscopic Morphology of Protoplasm, Amsterdam Elsevier. 1953

[7] Michelson A A. *Philos Mag*, 1911, **21**: 219

[8] Mauguin C. *Bull Soc Fr Mineral*, 1911, **34**: 6, 71

[9] de Vries M. *Acta Crystallogr*, 1951, **4**: 219

[10] Frank F C. *Discuss Faraday Soc*, 1958, **25**: 19

[11] Helfrich W. *Z Naturforsch*, 1973, **28**: C693

[12] Deuling H J, Helfrich W. *J Phys(Paris)*, 1976, **37**: 1335

[13] Ou - Yang Zhong - can, Helfrich W. *Phys Rev Lett*, 1987, **59**: 2486; *Phys Rev A*, 1989, **39**: 5280

[14] Naito H, Okuda M, Ou - Yang Zhong - can. *Phys Rev E*, 1993, **48**: 2304; Tanford C. *Proc Natl Acad Sci USA*, 1979, **76**: 3318

[15] Ou - Yang Zhong - can. *Phys Rev A*, 1990, **41**: 4517

[16] Mutz M, Bensimon D. *Phys Rev A*, 1991, **43**: 4525

[17] Rudolph A S, Ratna B R, Kahn B. *Nature*, 1991, **352**: 52

[18] Nakashima N, *et al. Chem Lett*, 1984, **1984**: 1709

[19] Yamada K, *et al. Chem Lett*, 1984, **1984**: 1713

[20] de Gennes P G. *C R Acad Sc Paris*, 1987, **304** (II): 259
[21] Lubensky T C, Prost J. *J Phys* II, 1992, **2**: 371
[22] Helfrich W, Prost J. *Phys Rev A*, 1988, **38**: 3065
[23] Ou-Yang Zhong-can, Liu Jixing. *Phys Rev Lett*, 1990, **65**: 1679; *Phys Rev A*, 1991, **43**: 6826
[24] Schnur J M. *Science*, 1993, **262**: 1669; Schnur J M, *et al*. *Science*, 1994, **264**: 945
[25] Selinger J, Schnur J M. *Phys Rev Lett*, 1993, **71**: 4091
[26] Chung D S, Benedek G, Konokoff F M, Donovan M. *Proc Natl Acad Sci USA*, 1993, **90**: 11341

(欧阳钟灿)

§4.3 轴对称类脂双层膜泡形状方程

相当长时间以来,对类脂双层膜泡轴对称形状的研究文献中,存在着三个不同的形状方程。但这三个方程是否都适于讨论轴对称膜泡形状问题,它们之间关系如何,长期未引起人们注意。本文从导出这三个方程所采用的独立变分函数的不同,系统地讨论了此问题,得出明确的结论。结论表明,常用的 Helfrich 方程可以用来研究轴对称膜泡,但由于它仅是普遍轴对称形状方程的一个特例,所给出的解有很大局限性。文献中常见的另一个方程,所谓 Peterson 方程,则由于推导该方程时所用的独立变分函数不能表示膜泡的轮廓变化,原则上是不正确的。仅在一种极为特殊的情况下,Peterson 方程退化为 Helfrich 方程,此时 Peterson 方程才可能给出正确的膜泡形状。

4.3.1 问题的提出

根据目前公认的关于生物膜结构的 Singer - Nicolson 模型[1],即生物膜的"流体镶嵌"模型,生物膜的主要成分为脂质、蛋白质和糖类。脂质是双亲分子,与水混合时,形成亲水头向水,厌水尾避水的类脂双层膜。蛋白质分子镶嵌在类脂双层膜的表面或贯穿其中,像"冰山"一样在流体态的类脂膜中浮动。大量的生物化学实验证明,流体镶嵌模型体现了生物膜的基本特征,虽然仍需改进,但不失为一个成功的模型。

在实验室开展生物膜的功能与结构的研究,主要用人工制备的双层脂膜。特别在生物膜的流体镶嵌模型被普遍接受,以及发展了在双层脂膜中镶入生物活性分子的技术后,人工脂膜的研究更为人们所重视。总结

1987 年以前所作的实验，张志鸿、刘文龙在他们所编著的《膜生物物理学》一书中给出了一张对照表，对人工类脂双层膜同生物膜的某些物理性质作了对比[2]。这里我们把这张表照抄下来，以便说明用类脂双层膜究竟可以在多大程度上作为真实生物膜的模型。

表 4.3.1 人工类脂双层膜与生物膜某些物理参数的对比

性 质	生 物 膜	类脂双层膜
厚度(10^{-10}m)	40～130	40～90
电容(μF/cm^2)	0.5～1.3	0.3～1.3
电阻(Ωcm^2)	10^2～10^5	10^3～10^9
击穿电压(mV)	100	100～550
静息电位差(mV)	10～88	0～140
折射率	约 1.6	1.37～1.66
界面张力(10^{-7}J/cm^2)	0.03～3.0	0.2～6.0
水透性(10^{-4}cm/s)	0.25～400	8.50
兴奋性	有	可以观察到
离子选择性	有	可以观察到
光激发	有	可以观察到

从上表不难看出，类脂双层膜在几何学、力学、电学、光学及某些生化性质等参数方面确实与真实生物膜十分近似，从而有理由认为对类脂双层膜的物理性质研究可以加深对生物膜物理性质的理解。

本节所要讨论的主题是关于在研究类脂双层膜泡形状时，基于 Helfrich 自发曲率弹性自由能导出的几个轴对称封闭膜泡方程究竟有什么关系这个有争议的问题。为了把问题的来龙去脉交代清楚，先谈一点历史。

1973 年，W. Helfrich 从类脂双层膜与丝状液晶的相似性出发，考虑到红细胞在形变时保持体积和表面积基本不变的实验事实，利用液晶弹性理论，提出了类脂双层膜泡的总弹性自由能表达式[3]：

$$F = \frac{k_c}{2}\oint(c_1 + c_2 - c_0)^2\mathrm{d}A + \lambda\int\mathrm{d}A + \Delta p\int\mathrm{d}V, \qquad (4.3.1)$$

其中 k_c，c_1，c_2，c_0，λ，Δp 分别表示膜的弯曲弹性系数，膜面的两个主曲率，膜两侧不对称的自发曲率，膜的表面张力系数和膜泡内外的压力差。类脂双层膜泡的平衡形状对应于表达式(4.3.1)取极小。由于(4.3.1)式中第一项曲率能中引入了自发曲率 c_0，(4.3.1)式又被称为 Helfrich 自发

曲率弹性自由能公式，简称 Helfrich 自由能公式。

1976 年，Helfrich 及其合作者从(4.3.1)式的轴对称表示

$$F_{\mathrm{H}} = 2\pi k_c \int L(\psi, \psi_x, x)\mathrm{d}x \tag{4.3.2}$$

出发，其中

$$L(\psi(x), \psi_x, x) = \left(\cos\psi \frac{\mathrm{d}\psi}{\mathrm{d}x} + \frac{\sin\psi}{x} - c_0\right)^2 \frac{x}{2\cos\psi} + \frac{\lambda x}{k_c \cos\psi} + \frac{\Delta p x^2 \sin\psi}{2k_c \cos\psi}。 \tag{4.3.3}$$

通过对膜泡经向和纬向主曲率作变分使 F_H 取极小，导出了第一个轴对称膜泡形状方程[4]：

$$\mathscr{H} \equiv \cos^2\psi \frac{\mathrm{d}^2\psi}{\mathrm{d}x^2} - \frac{\sin\psi\cos\psi}{2}\left(\frac{\mathrm{d}\psi}{\mathrm{d}x}\right)^2 + \frac{\cos^2\psi}{x}\frac{\mathrm{d}\psi}{\mathrm{d}x} - \frac{\sin 2\psi}{2x^2} - \frac{\Delta p x}{2k_c\cos\psi} - \frac{\lambda\sin\psi}{k_c\cos\psi} - \frac{\sin\psi}{2\cos\psi}\left(\frac{\sin\psi}{x} - c_0\right)^2 = 0。 \tag{4.3.4}$$

在(4.3.3)、(4.3.4)两式中，几何变量 x，ψ 的意义由图 4.3.1 表示，为了简化书写，将 $\mathrm{d}\psi/\mathrm{d}x$ 写为 ψ_x。为了方便，今后我们称方程(4.3.4)为 Helfrich 形状方程，它曾在轴对称膜泡研究中起过很大作用，Deuling 与 Helfrich 曾从这个方程求数值解得到人类红血球细胞形状及其各种变形[5]，苗玲等人也用此方程讨论过膜泡的长芽现象[6]。

1985 年，Peterson 先将对称膜泡轮廓曲线用曲线长度 s 参数化，从同样的自由能形式(4.3.2)出发，导出一组联立一阶微分方程组形式的轴对称形状方程[7]。此后，此组方程经常为 Seifert，Lipowsky 等人用来研究轴对称膜泡形状以及其相图[8,9]。如果以 $\psi(x)$ 为变量，这组联立方程可被表示为三阶偏微分方程：

$$\begin{aligned}\mathscr{D} \equiv{} & \cos^3\psi \frac{\mathrm{d}^3\psi}{\mathrm{d}x^3} - \cos^2\psi\left(3\sin\psi + \frac{1}{\sin\psi}\right)\frac{\mathrm{d}^2\psi}{\mathrm{d}x^2}\frac{\mathrm{d}\psi}{\mathrm{d}x} \\ & + \frac{2\cos^3\psi}{x}\frac{\mathrm{d}^2\psi}{\mathrm{d}x^2} + \cos\psi\sin^2\psi\left(\frac{\mathrm{d}\psi}{\mathrm{d}x}\right)^3 - \frac{(2+5\sin^2\psi)\cos^2\psi}{2x\sin\psi} \\ & \cdot\left(\frac{\mathrm{d}\psi}{\mathrm{d}x}\right)^2 + \left(\frac{\Delta p x}{2k_c\sin\psi} + \frac{c_0\sin\psi}{x} + \frac{\sin^2\psi}{x^2}\right)\cos\psi\frac{\mathrm{d}\psi}{\mathrm{d}x} \\ & - \left[\frac{\Delta p}{k_c} + \left(\frac{\lambda}{k_c} + \frac{c_0^2}{2}\right)\frac{\sin\psi}{x} - \frac{\sin\psi(1+\cos^2\psi)}{2x^3}\right] = 0。\end{aligned} \tag{4.3.5}$$

方程(4.3.5)称为 Peterson 方程。

1977 年,Jenkins 使用与(4.3.1)式略有差别的自由能表示式导出不含自发曲率 c_0 的描写任意形状双层膜泡的形状方程,并在文末的附注中对如果包含 c_0 时的方程形式作了简短说明[10]。1987 年,欧阳钟灿和 Helfrich 才直接从自由能表示(4.3.1)出发,导出了直接含自发曲率 c_0 的描写任意膜泡形状的普遍形状方程[11]:

$$2k_c\nabla^2H + k_c(2H + c_0)(2H^2 - 2K - c_0H) - 2\lambda H + \Delta p = 0, \tag{4.3.6}$$

其中 $H=(c_1+c_2)/2$ 为平均曲率,$K=c_1c_2$ 为高斯曲率,∇^2 为拉普拉斯-贝尔特拉米算符,他们利用此方程详尽地研究了类球膜泡的稳定性。此后欧阳钟灿还首先从方程(4.3.6)出发,得出过大小半径比为 $\sqrt{2}:1$ 的环解,并成功地为实验所证实[12]。

方程(4.3.6)可以描写任何形状的膜泡,不像 Helfrich 形状方程(4.3.4)和 Peterson 方程(4.3.5)那样只限于描写轴对称膜泡。与(4.3.6)相应的轴对称形状方程是什么样的呢?郑伟谋用 $\sin\psi$ 的幂级数展开系数表示给出过方程(4.3.6)的轴对称形式[13],后来,胡建国和欧阳钟灿将方程(4.3.6)轴对称化,用更紧凑的方式将其写出[14],得到:

$$\begin{aligned}\mathscr{G} \equiv & \cos^3\psi\frac{\mathrm{d}^3\psi}{\mathrm{d}x^3} - 4\sin\psi\cos^2\psi\frac{\mathrm{d}^2\psi}{\mathrm{d}x^2}\frac{\mathrm{d}\psi}{\mathrm{d}x} + \frac{2\cos^3\psi}{x}\frac{\mathrm{d}^2\psi}{\mathrm{d}x^3} \\ & + \cos\psi\left(1 - \frac{3}{2}\cos^2\psi\right)\left(\frac{\mathrm{d}\psi}{\mathrm{d}x}\right)^2 \\ & - \frac{7\sin\psi\cos^2\psi}{2x}\left(\frac{\mathrm{d}\psi}{\mathrm{d}x}\right)^2 - \left(\frac{c_0^2}{2} + \frac{\lambda}{k_c}\right. \\ & \left. - \frac{2c_0\sin\psi}{x} - \frac{1-3\cos^2\psi}{2x^2}\right)\cos\psi\frac{\mathrm{d}\psi}{\mathrm{d}x} \\ & - \left[\frac{\Delta p}{k_c} + \left(\frac{\lambda}{k_c} + \frac{c_0^2}{2}\right)\frac{\sin\psi}{x} - \frac{\sin\psi(1+\cos^2\psi)}{2x^3}\right] \\ = & 0。\end{aligned} \tag{4.3.7}$$

因为方程(4.3.7)是由普遍形状方程(4.3.6)轴对称化得到的,我们今后称之为普遍轴对称形状方程。

现在我们看到,描述轴对称膜泡的形状方程有三个,即 Helfrich 形状方程(4.3.4)、Peterson 方程(4.3.5)和普遍轴对称形状方程(4.3.7)。它

们在形式上有差别,Helfrich 形状方程是二阶方程,而 Peterson 方程和普遍轴对称形状方程则是三阶的。此外,推导它们的过程也不同,方程(4.3.4)和(4.3.5)是通过将轴对称化的自由能公式(4.3.2)变分而得,而方程(4.3.7)则是将普遍自由能形式(4.3.1)变分所得方程(4.3.6)轴对称化后得到的。人们自然会问,它们究竟是否等价?相互之间有什么关系?但长期以来,尽管不同作者使用不同的方程研究膜泡形状,这些问题并没有引起人们的特别注意。

郑伟谋虽然在以级数展开法求方程(4.3.6)的轴对称解时发现了 Helfirch 方程与普遍轴对称方程并不完全相同,但未作进一步讨论[13]。胡建国和欧阳钟灿首先明确指出这三个方程的不同[14]。他们从三个方程所能给出的解析解的异同、以及方程推导方法的不同对这三个不同形式的轴对称方程进行了比较,得出 Helfrich 方程只能给出正确的球泡解,Peterson 方程可以给出正确的球泡和柱形解,唯有普遍轴对称方程才能给出正确的球形、柱形和环形解的结论。至于三个方程为什么不同,他们当时的看法是因为对膜泡轮廓曲线参数化不同而造成的。此外,他们还认为从轴对称自由能公式推导普遍轴对称方程必须取廓线的法向变分,他们也确实通过对方程(4.3.2)取法向变分得到了方程(4.3.7)。

以上结论虽然正确指出了三个方程是不同的,但并未明确三个方程间的具体关系。三个方程不同,是否因为推导过程中膜泡轮廓曲线参数化不同所造成也值得商榷。另外,由轴对称自由能表示方程(4.3.2)直接导出普遍轴对称方程(4.3.7),是否一定要求在特殊方向取变分,也待讨论。

本节基于我们最近的研究结果[15,16],从变分法的基本原理出发,澄清以下几个问题:

① 普遍轴对称形状方程是否可以用一般变分法从轴对称自由能表示式(4.3.2)直接得出?它的推导与廓线如何参数化是否有关?是否一定要在特殊方向取变分?

② Helfrich 形状方程是不是一个正确的形状方程?它与普遍轴对称方程的关系如何?

③ Peterson 方程是正确的还是错误的?如果它是错的,是什么原因引起的?

④ 究竟是什么原因造成三个方程的不同?

4.3.2 普遍轴对称形状方程可以用一般变分法导出

这里讨论前节中提出的第一个问题，普遍轴对称方程是否可以直接由式(4.3.2)用一般的变分法得出?为了回答这个问题,我们首先必须弄清楚,从自由能泛函(4.3.2)作变分得出普遍轴对称形状方程在数学上的正确提法。

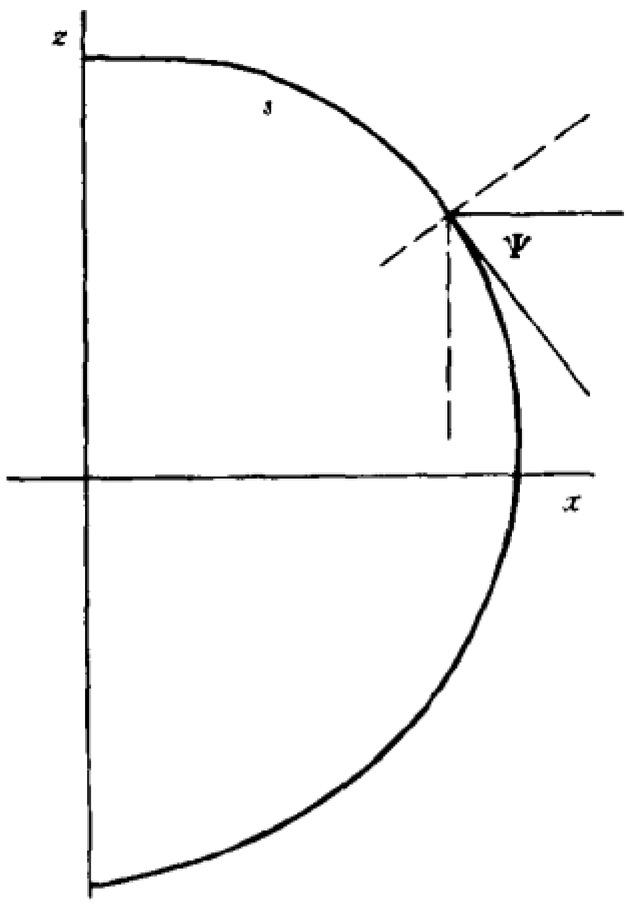


图 4.3.1 轴对称膜泡轮廓线示意图

其中对称轴为 z 轴, x 为廓线上任一点距对称轴的距离, ψ 为 x 轴与该点切线之间的夹角, s 为廓线弧长。

如图 4.3.1 所示,轴对称膜泡形状是由 x - z 平面上的廓线 $z(x)$ 描述的,这里 z 轴是对称轴。将自由能 F_H 取极小而得出形状方程在数学上是一个变分法问题,其正确提法是:在 z - x 平面上确定一条曲线 $z=z(x)$ 使得自由能泛函 $F_H[z(x)]$ 取极小。这里我们把方程(4.3.2)给出的自由能泛函 F_H 写成 $F_H[z(x)]$ 是为了强调 F_H 对廓线函数 $z(x)$ 的函数依赖关系。此一提法明确了当对方程(4.3.2)求变分时的独立变分函数。

我们要确定的是轴对称膜泡的廓线,因此令自由能泛函(4.3.2)相对 $z(x)$ 变化而变分

$$\delta_z F_{\rm H} = 2\pi k_c \int \frac{\delta L}{\delta \psi} \delta_z \psi \, {\rm d}x, \tag{4.3.8}$$

其中 $\delta_z\psi$ 是由于 δz 而引起的 ψ 的变分。由关系式

$$\frac{{\rm d}z}{{\rm d}x} = -\tan\psi, \tag{4.3.9}$$

可得

$$\delta_z\psi = -\cos^2\psi \frac{{\rm d}\delta z}{{\rm d}x}, \tag{4.3.10}$$

从而方程(4.3.8)成为

$$\delta_z F_{\mathrm{H}} = -2\pi k_c \int \frac{\delta L}{\delta \psi} \cos^2 \psi \frac{\mathrm{d}}{\mathrm{d}x}(\delta z)\mathrm{d}x$$

$$= 2\pi k_c \int \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\delta L}{\delta \psi} \cos^2 \psi \right) \delta z\, \mathrm{d}x, \qquad (4.3.11)$$

由此得出形状方程为

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\delta L}{\delta \psi} \cos^2 \psi \right) = 0, \qquad (4.3.12)$$

将 L 的具体表达式(4.3.3)代入(4.3.12)式,再将各项除以公因子 $1/x$,不难验证所得到的方程正好就是普遍轴对称形状方程(4.3.7)。

以上推导表明,普遍轴对称形状方程的确可以用一般的变分法从自由能泛函(4.3.2)导出,并不需要在特殊方向取变分。

下一个问题是,是否会因廓线的参数化不同而推导出不同的轴对称形状方程来?假定膜泡廓线以参数化方式$\{x=x(t),\ z=z(t)\}$表示,这里 t 为任意参数。此时,与(4.3.9)式相对应的几何关系式成为

$$\frac{z_t}{x_t} = -\tan\psi, \qquad (4.3.13)$$

其中 $z_t=\mathrm{d}z/\mathrm{d}t$, $x_t=\mathrm{d}x/\mathrm{d}t$。引入拉格朗日乘子将约束条件(4.3.13)加到自由能泛函数(4.3.2)内,则可将 $z(t)$, $\psi(t)$, $x(t)$看作独立函数,新的自由能泛函可写作:

$$\widetilde{F} = 2\pi k_c \left[\int \widetilde{L}(\psi(t), \psi_t, x_t, x(t)) x_t \mathrm{d}t + \int \gamma \left(\frac{z_t}{x_t} + \tan\psi \right) \mathrm{d}t \right], \qquad (4.3.14)$$

其中 γ 为局域拉氏乘子,且

$$\widetilde{L}(\psi(t), \psi_t, x_t, x(t)) \equiv L(\psi, \psi_t/x_t, x)。 \qquad (4.3.15)$$

自由能泛函 $\widetilde{F}$ 对应于函数 $z(t)$, $\psi(t)$, $x(t)$的欧拉-拉格朗日方程分别为:

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\gamma}{x_t} \right) = 0, \qquad (4.3.16)$$

$$\frac{\delta(\widetilde{L} x_t)}{\delta \psi} + \frac{\gamma}{\cos^2\psi} = 0, \qquad (4.3.17)$$

$$x_t \frac{\partial \widetilde{L}}{\partial x} - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial \widetilde{L}}{\partial x_t} x_t - \widetilde{L} \right) - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\gamma \tan\psi}{x_t} \right) = 0。 \qquad (4.3.18)$$

将方程(4.3.16)代入(4.3.17)得：

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\cos^2\psi\,\frac{\delta(\widetilde{L}x_t)}{x_t\delta\psi}\right]=\frac{\mathrm{d}}{\mathrm{d}t}\left(\cos^2\psi\,\frac{\delta L}{\delta\psi}\right)=0,\qquad(4.3.19)$$

这正好是方程(4.3.12)，也就是普遍轴对称形状方程(4.3.7)。

注意我们在导出方程(4.3.19)时未用到方程(4.3.18)，因此还应该检查(4.3.18)和(4.3.19)是否矛盾。我们在文献[14]所作的检验证明，这两个方程是相容的。

以上推导清楚地表明，普遍轴对称形状方程的确可以从轴对称自由能表示式(4.3.2)用一般的变分法直接推导出来。而且，轴对称膜泡廓线的参数化对方程推导毫无影响，不管廓线是如何参数化的，只要将独立变分函数选为廓线，都会导出同样的形状方程。

4.3.3 Helfrich 形状方程的局限性

回过头我们来讨论 4.3.1 节末提出的第二个问题，即最早用来研究轴对称膜泡形状的 Helfrich 形状方程是否正确以及它与普遍轴对称形状方程的关系。

注意前文给出的普遍轴对称形状方程(4.3.12)或(4.3.19)都是全微分形式，这表明，普遍轴对称形状方程(4.3.7)虽然看起来非常复杂，而且是高度非线性的，其实这个方程存在一次积分。将方程(4.3.12)或(4.3.19)对 x 或 t 积分，立即得到：

$$\frac{\delta L}{\delta\psi}\cos^2\psi=C,\qquad(4.3.20)$$

其中 C 为积分常数。方程(4.3.20)是方程(4.3.7)的一次积分形式，因此使用起来应当比(4.3.7)更方便。如果令(4.3.20)中的积分常数为零，立即可得：

$$\frac{\delta L}{\delta\psi}\equiv\frac{\partial L}{\partial\psi}-\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial L}{\partial\psi_x}\right)=0,\qquad(4.3.21)$$

此方程两端如乘一 $\cos\psi/x$ 因子，再将 L 的具体表达式代入，马上给出 Helfrich 形状方程(4.3.4)。

这说明，Helfrich方程实际上是普遍轴对称形状方程(4.3.7)的一次积分的积分常数取作零时的特殊形式。事实上，在文献[15,16]中，我们曾通过对方程(4.3.4)和方程(4.3.7)的代数结构的讨论，给出过普遍轴对称形状方程与 Helfrich 形状方程之间的这一关系，即

$$\mathscr{G}=\frac{1}{x}\frac{\mathrm{d}}{\mathrm{d}x}(x\cos\psi\mathscr{H}),\tag{4.3.22}$$

和

$$x\cos\psi\mathscr{H}=C,\tag{4.3.23}$$

注意方程(4.3.20)与方程(4.3.23)是同一方程的不同写法。

这样，我们对Helfrich形状方程的正确性和局限性就全都清楚了。首先，这个方程是正确的，可以用它来研究某些轴对称膜泡形状。其次，这个方程是有局限的轴对称膜泡形状方程，它只能用来研究方程(4.3.21)或(4.3.23)中积分常数为零的那一部分膜泡的形状。因此，它只能给出部分轴对称膜泡的形状的解，而不能像普遍轴对称膜泡形状方程(4.3.7)或(4.3.21)、(4.3.23)那样，给出所有可能的轴对称膜泡形状的解。

那么，为什么从同样的自由能泛函(4.3.2)出发，会导出不同的形状方程呢？具体地说，Helfrich 等人当时推导方程时，出了什么漏洞使得他们导出一个并不完整的形状方程呢？

看一下他们的推导，原因就很清楚了。他们在对轴对称形式自由能泛函作变分时，取的独立变分函数不是膜泡形状的廓线 $z(x)$，而是廓线上一点的切线与 x 轴之间的夹角 ψ。这样，与自由能泛函(4.3.2)对应的欧拉-拉格朗日方程便是：

$$\frac{\delta L}{\delta\psi}\equiv\frac{\partial L}{\partial\psi}-\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial L}{\partial\psi_x}\right)=0。$$

如前所述，将此式乘因子 $\cos\psi/x$，即为 Helfrich 形状方程(4.3.4)。我们在本节开头曾对通过将自由能泛函(4.3.2)极小化求膜泡形状方程的正确数学提法作过陈述，指出此问题是要在 $z-x$ 平面上确定一条膜泡廓线 $z=z(x)$ 使得(4.3.2)式取极小。所以，作变分时独立变分函数只能是廓线本身。夹角 ψ 与廓线虽有关系，但不能唯一地决定廓线。Helfrich 等人推导形状方程时选它为独立变分函数，必然导致推出的方程与普遍轴对称形状方程不同。

4.3.4 Peterson 方程原则上与普遍轴对称形状方程不相容

4.3.3 讨论了 Helfrich 形状方程与普遍轴对称形状方程的关系，这里我们考察 Peterson 方程与普遍轴对称形状方程的关系，并讨论其正确

性，从而对 4.3.1 未提出的第三个问题作出答复。

首先来看 Peterson 方程与普遍轴对称形状方程间的关系。从方程(4.3.5)中减去方程(4.3.22)，我们得到：

$$\mathscr{P}-\mathscr{G}=-\mathscr{H}\frac{\cos^2\psi}{\sin\psi}\frac{\mathrm{d}\psi}{\mathrm{d}x}。\tag{4.3.24}$$

因此，

$$\begin{aligned}\mathscr{P}&=\mathscr{G}-\mathscr{H}\frac{\cos^2\psi}{\sin\psi}\frac{\mathrm{d}\psi}{\mathrm{d}x}\\&=\frac{1}{x}\frac{\mathrm{d}}{\mathrm{d}x}(x\cos\psi\mathscr{H})-\mathscr{H}\cos\psi\frac{\cos\psi}{\sin\psi}\frac{\mathrm{d}\psi}{\mathrm{d}x}\\&=\frac{\sin\psi}{x}\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{x\cos\psi\mathscr{H}}{\sin\psi}\right)。\end{aligned}\tag{4.3.25}$$

将等式(4.3.25)的右端代入方程(4.3.5)，马上可以看出 Peterson 方程是可积的，它实际上存在一个一次积分

$$\frac{x\cos\psi\mathscr{H}}{\sin\psi}=C',\tag{4.3.26}$$

其中，C'是积分常数。对比方程(4.3.23)和(4.3.26)，可以立即得出以下两点结论：

① 当(4.3.26)式的积分常数为零时，Peterson 形状方程等价于 Helfrich 形状方程。

② Peterson 方程原则上与普遍轴对称形状方程不相容，这可由(4.3.26)式右端存在分母 $\sin\psi$ 明显地表现出来。只有在两种特殊情况下，普遍轴对称形状方程与 Peterson 方程才可能有共同解。一种情况是当$C=C'=0$时，此时二者均退化为 Helfrich 方程；另一种情况是它们都有柱形解 $\psi=\pi/2$，此时，$\sin\psi=1$，退化为常数。在其他情况下，两者没有共同解。

由此可见，使用 Peterson 方程研究轴对称膜泡形状必须特别谨慎，除了它退到 Helfrich 形状方程时给出的解以外，其积分常数不为零的解与实际膜泡形状都毫无关系。

使得 Peterson 方程出毛病的原因是什么？让我们看一下 Peterson，Seifert，Lipowsky 等人[7,8]推导这个方程的过程。

他们首先用弧长将膜泡廓线参数化，并使用约束条件

$$x_s = \cos\psi, \tag{4.3.27}$$

把约束条件乘上拉格朗日乘子加入自由能泛函(4.3.2)中，得

$$F' = 2\pi k_c\left[\int L\left(\psi(s),\ \frac{\psi_s}{x_s},\ x(s)\right)x_s\,\mathrm{d}s + \int \gamma(x_s - \cos\psi)\mathrm{d}s\right]。 \tag{4.3.28}$$

对 F' 变分，得出对应于 $\psi(s)$，$x(s)$ 的欧拉-拉格朗日方程分别为：

$$\frac{\delta(Lx_s)}{\delta\psi} + \gamma\sin\psi = 0, \tag{4.3.29}$$

和

$$\frac{\delta(Lx_s)}{\delta x} - \frac{\mathrm{d}\gamma}{\mathrm{d}s} = 0。 \tag{4.3.30}$$

在一些已发表的论文中，Peterson，Seifert 等使用(4.3.29)～(4.3.30)为研究轴对称膜泡形状的基本方程组[8]，由(4.3.29)和(4.3.30)中消去 γ，即得到等价于这组联立方程的高阶方程：

$$\frac{\mathrm{d}}{\mathrm{d}s}\left[\frac{\delta(Lx_s)}{\delta\psi}\,\frac{1}{\sin\psi}\right] + \frac{\delta(Lx_s)}{\delta x} = 0, \tag{4.3.31}$$

将 L 的显式代入后，不难看出(4.3.31)就是方程(4.3.5)。

这一推导的问题在于他们对自由能泛函(4.3.2)变分时，实际上仍然没有选取膜泡廓线作为独立变分函数。因为，推导中对廓线以弧长 s 作参数化，此时廓线变化本来意味着两个函数 $x=x(s)$，$z=z(s)$ 都可独立变化。然而上面的推导仅允许廓线的一个坐标函数 $x(s)$ 独立变化，而不考虑另一个坐标函数 $z(s)$ 独立变化，这说明他们并没有把廓线函数作为泛函变分的独立变分函数。

为了更清楚地说明以上推导出来的方程(4.3.31)只不过是泛函(4.3.2)对 $x(s)$ 变分所得的欧拉-拉格朗日方程，下面给出一个直接计算。

参数化后的自由能泛函(4.3.2)为：

$$F_{\mathrm{H}} = 2\pi k_c\int L(\psi,\ \psi_x,\ x)x_s\,\mathrm{d}s, \tag{4.3.32}$$

该式关于 $x(s)$ 变化的变分为：

$$\delta F_{\mathrm{H}} = 2\pi k_c\int \delta L(\psi(s),\ \psi_s/x_s,\ x(s))x_s\,\mathrm{d}s$$

$$=2\pi k_c\int\left(\frac{\delta(Lx_s)}{\delta\psi}\delta\psi+\frac{\delta(Lx_s)}{\delta x}\delta x\right)\mathrm{d}s$$

$$=2\pi k_c\int\left[-\frac{\delta(Lx_s)}{\delta\psi}\left(\frac{1}{\sin\psi}\frac{\mathrm{d}\delta x}{\mathrm{d}s}\right)+\frac{\delta(Lx_s)}{\delta x}\delta x\right]\mathrm{d}s$$

$$=2\pi k_c\int\left[\frac{\mathrm{d}}{\mathrm{d}s}\left(\frac{\delta(Lx_s)}{\delta\psi}\frac{1}{\sin\psi}\right)+\frac{\delta(Lx_s)}{\delta x}\right]\delta x\mathrm{d}s。\qquad(4.3.33)$$

令 $\delta F_H=0$，自然得出方程(4.3.31)即方程(4.3.5)，所以它不是正确的形状方程。

通过以上各节的讨论，我们给出了关于三个轴对称膜泡形状方程之间的关系，也澄清了有关方程推导中的一些争论的问题。

讨论表明，普遍轴对称膜泡形状方程(4.3.7)是唯一正确的描写轴对称膜泡形状的方程。Helfrich 形状方程是普遍形状方程的特殊形式，它只能给出部分轴对称膜泡形状的解。而 Peterson 方程则原则上与普遍形状方程不相容，仅在特殊情况下退化为 Helfrich 形状方程。

讨论还表明，从轴对称形式的自由能泛函(4.3.2)推导形状方程时，关键是应把膜泡廓线选作独立变分函数，否则推导出来的方程便有问题，Helfrich 形状方程和 Peterson 方程的毛病都出在独立变分函数没有选对这一点上。

此外，我们还证明了，普遍膜泡形状方程存在一次积分，使用这个一次积分形式研究轴对称膜泡形状，应该比直接使用方程(4.3.7)更方便。

参 考 文 献

[1] Singer S J, Nicolson G L. *Science*, 1972, **175**: 720
[2] 张志鸿，刘文龙. 膜生物物理学. 北京：高等教育出版社，1987, 256
[3] Helfrich W. *Naturforsch Z*, 1973, C28: 693
[4] Deuling H, Helfrich W. *J Phys (Paris)*, 1976, **37**: 1335
[5] Deuling H, Helfrich W. *Biophys J*, 1976, **16**: 861
[6] Miao L, Fourcade B, Rao M, Wortis M, Zia R. *Phy Rev*, 1991, **A43**: 6843
[7] Peterson M. *J Appl Phys*, 1985, **57**: 1739
[8] Seifert U. *Phys Rev*, 1991, **A43**: 6803
[9] Seifert U, Berndl K, Lipowsky R. *Phys Rev*, 1991, **A44**: 1182
[10] Jenkins J. *Math JBiophys*, 1977, **4**: 149

[11] Ou-Yang Z C, Helfrich W. *Phys Rev Lett*, 1987, **59**: 2486; *Phys Rev*, 1989, **A39**: 5280
[12] Ou-Yang Z C. *Phys Rev*, 1990, **A41**: 4517; Mutz M, Bensimon B. *Phys Rev*, 1991, **A43**: 4525
[13] Zheng W M, Ou-Yang Z C. *Commun Theor Phys*, 1991, **15**: 505
[14] Hu J G, Ou-Yang Z C. *Phys Rev*, 1993, **E47**: 641
[15] Zheng W M, Liu J X. *Phys Rev*, 1993, **E48**: 2856; *Commun Theor Phys*, 1994, **21**: 369
[16] Zheng W M, Liu J X, Hu J G. *Chin Phys Lett*, 1994, **11**: 227

(刘寄星 郑伟谋)

第 5 章　生物数学模型

§5.1　免疫和肿瘤生长的非线性研究

免疫能对机体产生保护作用。当具有抗原性的物质进入机体后，免疫系统能识别"己"和"非己"，能排除或消灭抗原性的异物，这种消除可以是特异性的也可以是非特异性的，但特异性的更为重要。免疫系统也能被诱导得对于抗原性的异物处于不活化的状态，即免疫耐受。免疫系统还能监视机体自身细胞的癌变，并能杀死癌细胞，称抗癌免疫监视。

免疫系统由免疫器官、免疫细胞和免疫分子所组成。从骨髓分化出的淋巴系干细胞进入胸腺发育为各亚群 T 细胞，留在骨髓中的发育为 B 细胞。非 T 非 B 前体细胞则发育为 K 细胞和自然杀伤细胞(NK)等。在骨髓中增殖分化出的髓系干细胞将发育为巨噬细胞(Mφ)、中性粒细胞、嗜酸性粒细胞和嗜碱性粒细胞等。T 细胞和 B 细胞在全身血液进行循环和在外周淋巴组织中定居增殖，形成细胞免疫和体液免疫。巨噬细胞能吞噬抗原，又能与树突状细胞等呈递抗原。K 细胞和 NK 细胞则专门杀伤靶细胞。肥大细胞和嗜碱性粒细胞介导 IgE 抗体产生过敏反应。上述中性和嗜酸性粒细胞也在免疫中发挥作用。此外还有由细胞分泌出的多种细胞因子以及由巨噬细胞、肠上皮细胞和肝细胞等产生的多种多样的补体，与免疫细胞协同发挥作用，形成复杂的免疫系统。

免疫学在本世纪有很大发展。Burnet 于 1955—1957 年提出细胞克隆选择学说 (clonal selection theory)；1959 年 R. R. Porter 和 G. M. Edelman 发现免疫球蛋白分子结构；1961 年 J. F. Miller 发现胸腺的免疫功能；1966 年 H. N. Claman 等人报道 B 细胞与 T 细胞的相互作用；到 1974 年，著名免疫学家 N. K. Jerne 提出了独特型网络调节学说；20 世纪 70 至 80 年代，G. J. F. Köhler 和 C. Milstein 等人得到单克隆抗体，S. Tonegawa 提出免疫球蛋白基因结构。在近 40 年的免

疫学发展中，有两种学说占主导地位，即细胞克隆选择学说和独特型网络调节学说[1]。

所谓克隆系指细胞的无性繁殖系。克隆选择学说或称作细胞系选择学说，是近代免疫学的基础。它认为在个体内存在着识别各种抗原的许多免疫细胞系或称细胞克隆。各细胞系中的细胞通过其表面的受体特异地结合抗原。抗原起着选择活化的作用，促使带有相应受体的B细胞系的细胞活化增殖，发展为产生抗体的浆细胞并分泌抗体。据估计，关于B细胞的特异性和多样性的数量级可达10^7。每个B细胞表面上的抗体数量可达10^5的数量级。1970年G. Bell等曾运用数学模型研究克隆选择学说，对当时已知的免疫系统的工作作了相当全面的讨论[2,3]。自那时起，关于理论免疫学的讨论随着免疫学的发展而发展。例如近几年出现有关抗原呈递细胞(APC)与T细胞激活的讨论[4]，关于抗体高变区经基因编码从而以更高的亲和力为抗原所选择的过程[5]和关于对艾滋病毒(HIV)的免疫响应等，不胜枚举。

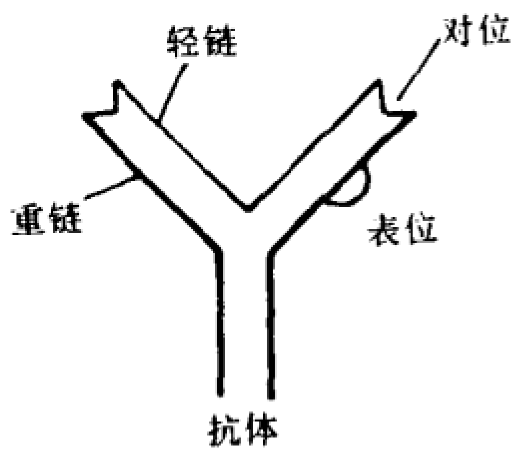


图 5.1.1 抗体示意图

它有两条重链和两条轻链，中间以二硫键联结，图示意地表示它既有识别其他抗体的对位，又有能被其他抗体识别的表位。

应当提到的是，近来相当一部分关于免疫的理论研究集中于独特型网络调节(idiotypic network regulation)。在免疫系统中，免疫细胞和它分泌的抗体上的对位(paratope)或者结合部位能识别和结合抗原的表位(epitope)或者抗原决定簇。值得注意的是，免疫细胞和抗体有双重作用，即具有特异性的免疫细胞和抗体上还存在具有抗原性的位点，称独特位(idiotope)或独特型决定簇(idiotypic determinant)，能被其他类型抗体的对位特异地识别。这后面的抗体的独特位又能被另种抗体的对位识别，依此类推。能识别某种独特型的称为抗独特型(anti-idiotope)。这样便形成多层次的独特型网络调节。体内的抗体即使在无抗原刺激下亦能因识别其他抗体而增殖，并且又因被另一些抗体识别而削弱，从而使得各种抗体数量在体内达到某种平衡。由于抗体同时也扮演了抗原的角色，因而独特型便被称作外部抗原的"内映象"。

各种独特型网络调节的数学模型已被提出来，例如形态空间模型

(shape space model)[6]。该模型在抗体与抗原识别和结合部位间存在非共价键的、基于静电力、氢键和范德华力等近距作用这一认识的基础上，用有限的 n 个参数描述结合部位的特征；或云，n 维空间的一点代表一抗体及其对位即结合部位，在其周围划出一小范围。如其他抗体的表位进入这一小区域则被抗体所识别。因该模型涉及时空，故演化方程表现为偏微分方程。另一种模型可称作字条模型。它视每一抗体包含表位 e 和对位 p，记作 (p,e)；设它们均由两种氨基酸的序列组成，分别用 0 和 1 表示，并随机排列。当 p_j 与 e_i 相遇时，不同数码相遇则相互匹配，相同数码相遇则不匹配，如图 5.1.2。相互能匹配的数码对越多，对位和表位的亲和力越强；相互能匹配的数码对少到一定程度则对位和表位不能结合[7]。利用这样的模型亦可观察在独特型网络调节中各种抗体数量的演化。最后，关于独特型网络调节的更多研究是运用非线性动力学方程：选择不同种

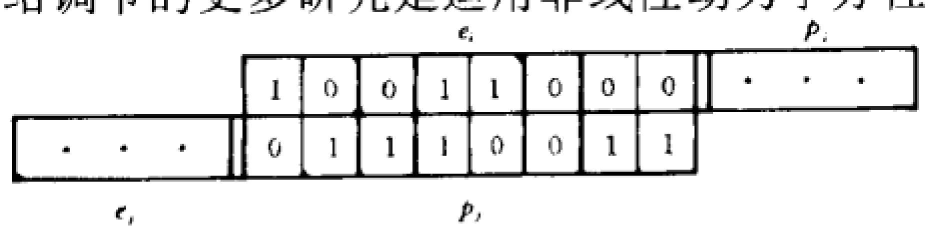


图 5.1.2　字条模型

(e_i,p_i)和(e_j,p_j)二字条表示两个不同的抗体，在本图中，有6对数码是互相匹配的。

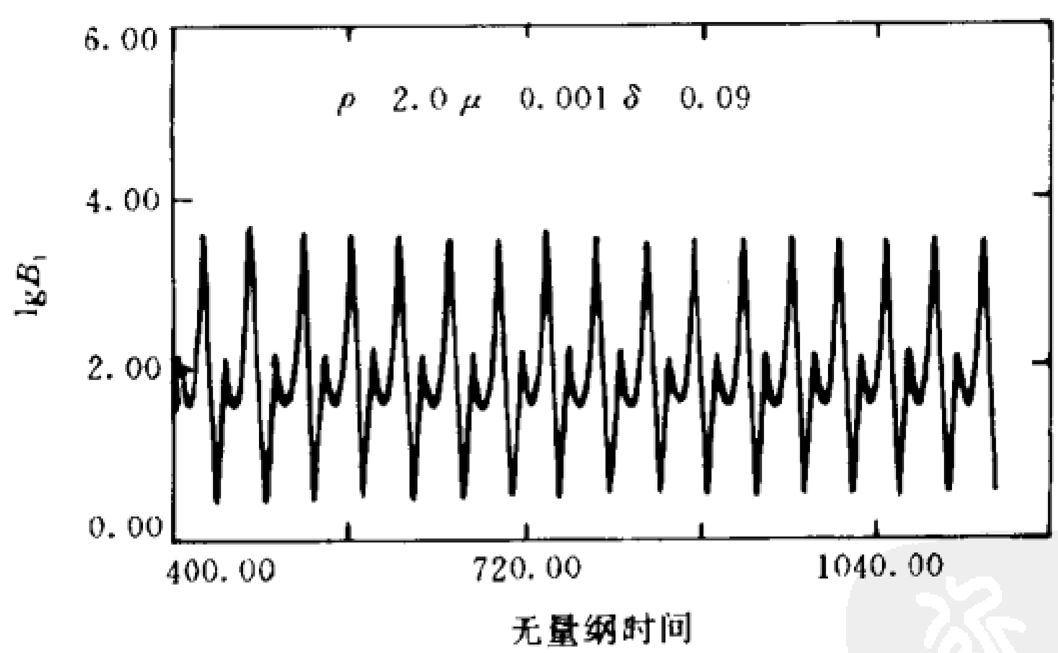


图 5.1.3　AB 模型关于抗体密度的周期解

B_1 代表 B 细胞浓度，横轴表示经重新标度的无量纲时间，ρ、μ 和 δ 则表示由 B 细胞的最大增殖率和死亡率、抗体和抗原形成的复合物的死亡率等组成的联合参数。(引自文献[19])

类的 B 细胞，它们分泌的抗体、因对位和表位相结合而形成的结合体等组分的浓度作为变量，以它们的产生率、消失率和结合率等作为参数，便可建立独特型网络调节的非线性模型。该模型可简化为仅含两种 B 细胞和两种抗体的 AB 模型。其分叉图与原模型的分叉图仅表现为在参数空间发生移动而并无本质的不同，这里的分叉属于 Hopf 分叉。越过分叉点后，定态失稳，进入周期振荡与混沌交替出现的参数空间[8,9~11]。目前，定态、混沌和周期解的生物学意义尚属于有待研究的问题。定态有平凡态，也有表示抗体密度处于高水平的定态，不必用图表示。图 5. 1. 3 表示周期解，而图 5. 1. 4 表示经倍周期分岔途径进入的混沌；此外，尚有阵发混沌发生。F. Varela 等人和 Lundkvist 等人均曾通过实验研究了在无外部抗原物质侵入情况下小鼠体内几种抗体数量随时间的演化。未与抗原接触的抗体叫做自然抗体。图

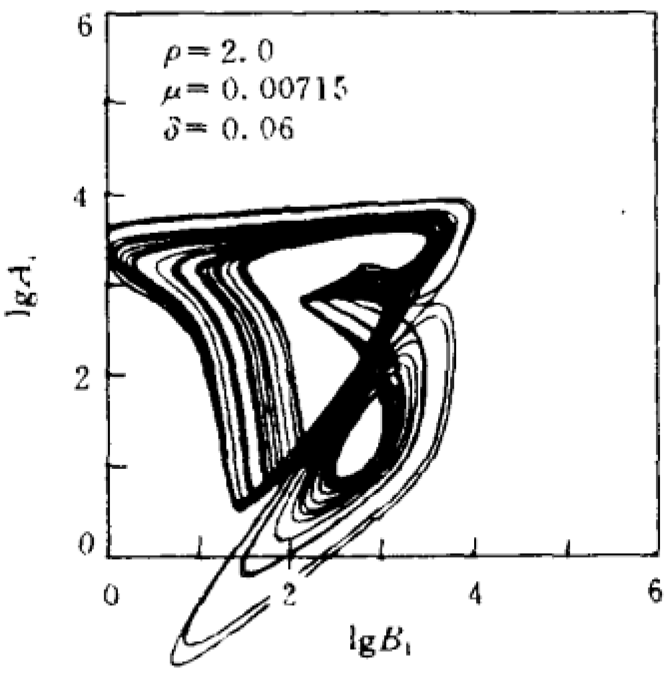


图 5.1.4 AB 模型关于抗体浓度的混沌行为

A_1、B_1 分别表示抗体和 B 细胞的浓度。ρ、μ和δ的含义与图5.1.3相同。(引自文献[9])

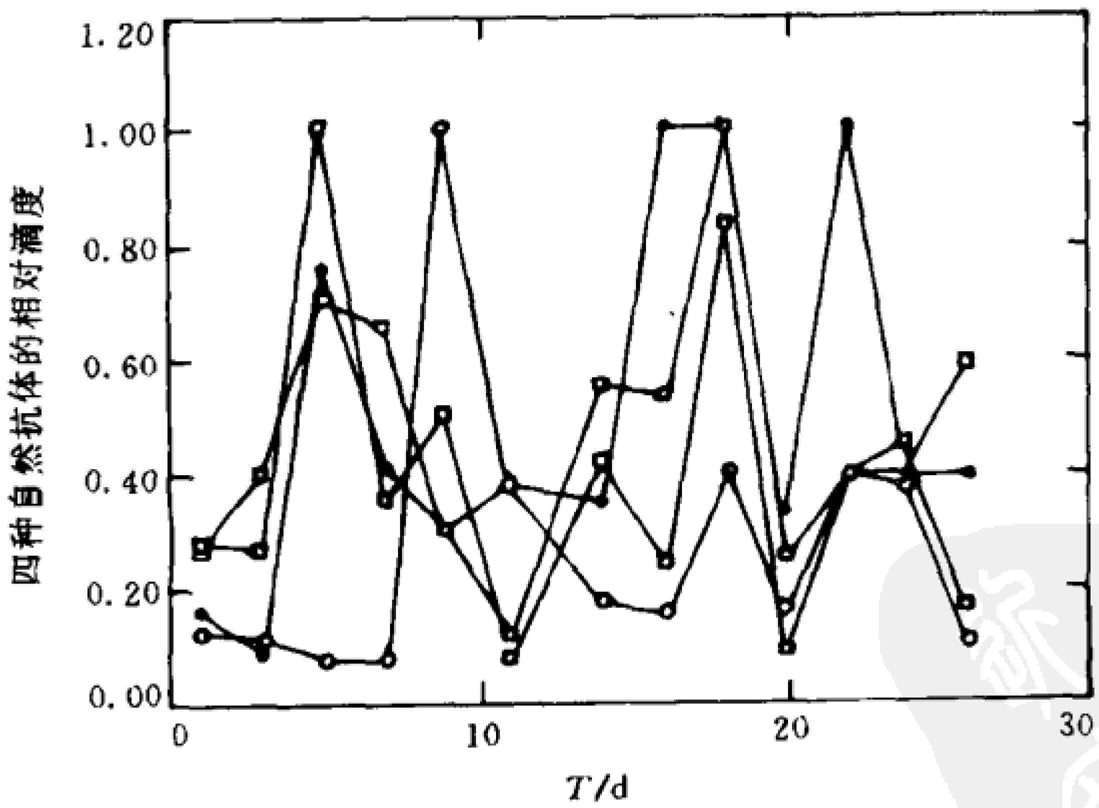


图 5.1.5 Varela 等人的实验结果

给出四种自然抗体相对滴度的时间演示——不规则的振荡行为。(引自文献[10])

5.1.5 表示 Varela 等人的结果[10]，它给出四种自然抗体的滴度或效价的相对值随时间的变化。这里的滴度或效价是实验免疫学中的名词，它反映血清中独特型的浓度。将其最大值归一，则得相对值随时间的演化。文章作者运用已知数据计算功率谱的方法研究相对滴度的演化是否具有混沌的特征。报道指出由于放血频率的限制，所得数据尚不足以给出充分的谱分析。

在用非线性理论研究免疫方面，免疫耐受也是大家关心的课题。较多的工作是运用独特型网络调节解释耐受，例如 G. Weisbuch 等人曾研究各种不同模式的独特型网络，并指出在某些网络中和在相当宽的参数范围内有可能形成免疫耐受。

除去独特型网络调节外，另一引起较多作者注意的问题为免疫记忆和免疫的二次响应。机体能消除入侵的抗原，当同类抗原第二次侵入机体时，抗体水平将更高，对抗原的消灭将更有力。这种对侵入抗原的记忆性是免疫系统的重要特征。人们对免疫记忆机制的认识尚不完全明了。大体有如下几种基本观点：第一种认为一部分 B 细胞接受抗原刺激后，转化为长寿的、静息的记忆 B 细胞。该过程不伴随抗体高变区氨基酸序列的突变，从而能对特异的抗原有更高的亲合力；它藏于体内，一旦抗原二次入侵，即被激活、增殖、分泌抗体并进行二次应答。第二种，认为免疫记忆并非源于特殊的记忆细胞，而是通过独特型网络调节来实现；即在抗原刺激下，网络从原始态转变至免疫态；具有抗原特异性的 B 细胞处于活化状态，并分泌大量抗体；即使抗原消灭后，免疫态能在抗原“内映象”的刺激下得以保持，以待抗原再一次的侵入。第三种，认为存在一种长寿细胞系。因外周淋巴器官有许多滤泡网状细胞(follicular dendritic cells，FDC)网络可长期储存抗原。小部分抗体抗原结合的复合物长期存在于 FDC 的表面。由于它的存在，使得一个长寿命的 B 细胞系得以维持，而单个 B 细胞并不必定是长寿的。抗原的再次侵入将因长寿命 B 细胞克隆的存在，而诱导出高水平的抗体以实现二次响应[12]。实际上，许多学者将上面谈到的不同观点结合起来解释免疫记忆的观点，例如既存在长寿命 B 细胞也存在长寿命克隆的观点，以及 B 记忆细胞与网络调节结合起来的观点等等，各种观点均有文章发表。然而，无论从理论到实验均尚未给出免疫记忆的完整描述。理论工作的一种方法是先从较少

因素出发描述现象的大概轮廓，然后引入更多因素以探索它们对现象的影响。既然最新实验结果报道长寿命B记忆细胞克隆的存在，便不妨以此为出发点从理论上探索免疫记忆。

该模型考虑了普通B细胞、记忆B细胞、抗体产生细胞即浆细胞、抗体和抗原诸组分，考虑了它们的产生和消灭以及相互作用，应用激活函数反映免疫细胞的激活程度，建立起关于免疫记忆的非线性动力学方程组，对于定态解和时间演化解进行了讨论。图5.1.6表示了抗原初次和再次入侵抗体时浓度变化情况，纵坐标和横坐标分别表示抗体无量纲浓度和时间。该图反映了二次响应的基本特征。理论结论还能显示诸种因素对二次响应的影响。例如，抗原第一次至第二次入侵的时间间隔 τ 有一临界值 τ_0，当 $\tau>\tau_0$ 时，模型才会表现出较强的二次应答。通过模型的研究还能给出记忆细胞的增殖率和分泌抗体的速率对二次应答的影响。通过长寿命B记忆细胞和长寿命B记忆细胞克隆的比较，还发现在二次免疫响应后，记忆B细胞的密度将长期保持高水平，这对于前一种观点是不合理的[13]，如图5.1.7所示。

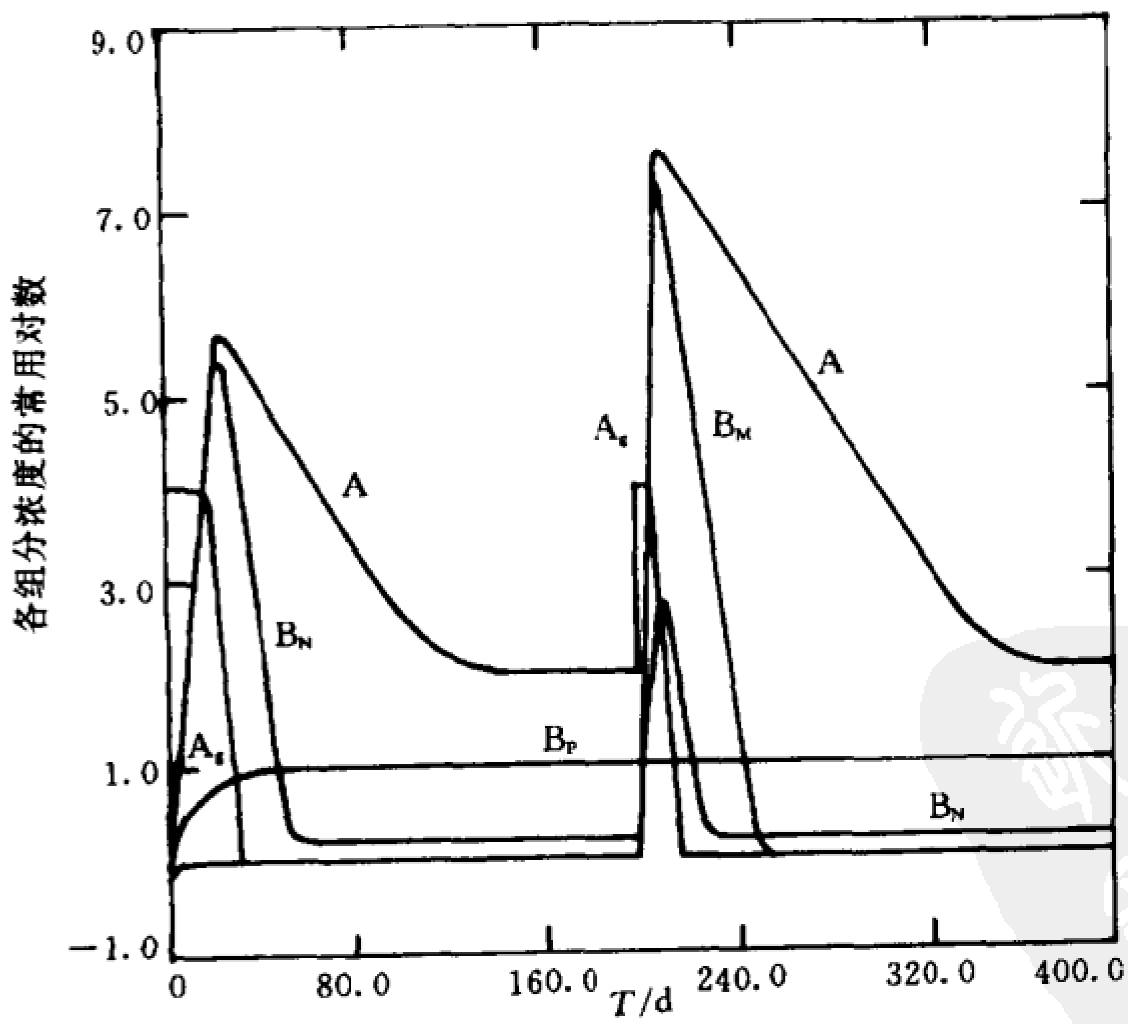


图 5.1.6 长寿命记忆克隆模型显示的二次响应

A_g、B_N、B_P、B_M 和 A 分别代表抗原、B细胞、浆细胞、记忆B细胞和抗体诸组分。(引自文献[13])

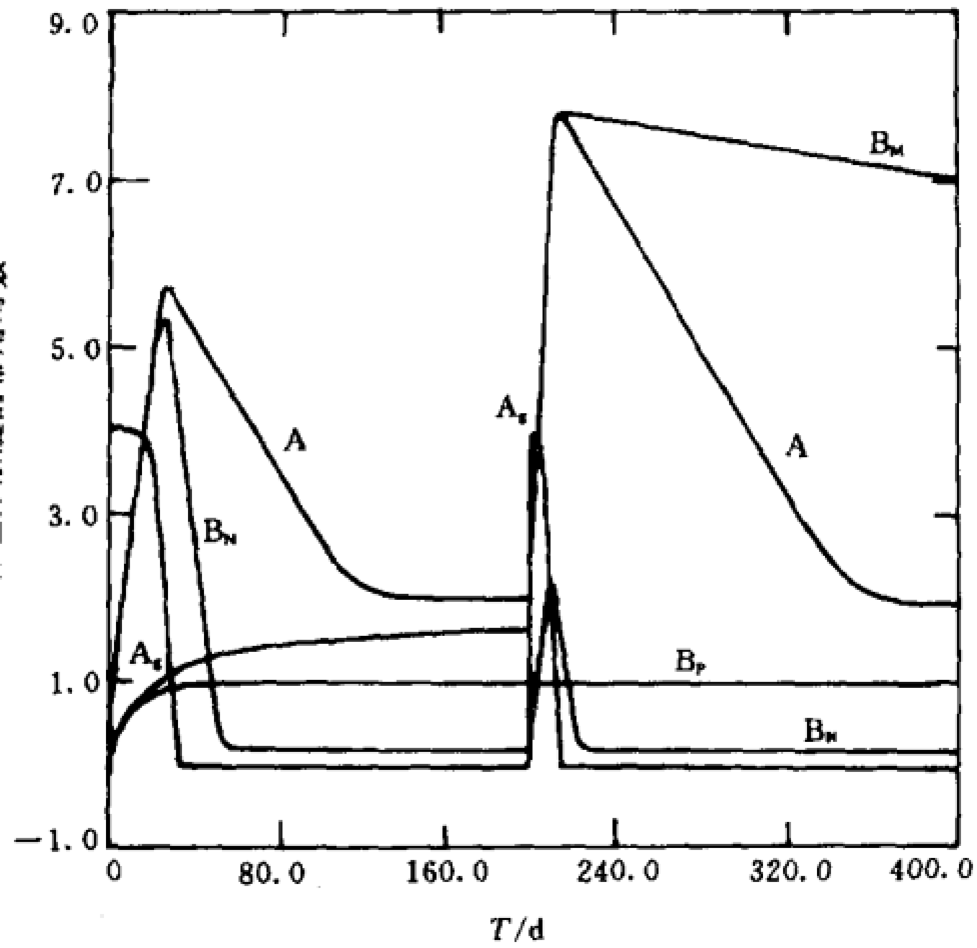


图 5.1.7　长寿命记忆细胞模型给出的二次响应

符号含义与图 5.1.6 相同。(引自文献[13])

在当前理论免疫学的研究中，T 细胞在胸腺微环境中的分化和发育也是大家关心的课题。在免疫监视下的肿瘤生长从来就是理论生物学关心的课题。现在回忆目前对此问题的基本认识，在肿瘤免疫中由 T 细胞介导的免疫应答起着重要作用。这种免疫应答也形成一复杂的网络，大体上说，T 细胞具有抗肿瘤的效应。自然杀伤细胞(NK)对肿瘤细胞有高度的细胞毒作用。K 细胞虽不能直接与肿瘤细胞结合，但借助于特异抗体的协同作用亦能激活 K 细胞，使之对肿瘤发生细胞毒作用。特别应当指出的是，巨噬细胞在抗肿瘤作用中起重要作用。它不仅能呈递肿瘤抗原，且能溶解肿瘤细胞。在 R. Lefever 的模型中考虑了肿瘤细胞的增殖，将对肿瘤的复杂的应答粗粒化为效应细胞对癌细胞的细胞毒作用。效应细胞对癌细胞的结合和溶解，则采用关于酶的 Michaelian 动力学。得出涉及癌细胞、效应细胞、效应细胞与癌细胞的结合体以及死亡的癌细胞四组分的偏微分方程组，它具有波前解。波前常出现于一类非线性反应扩散系统中。它在不同问题中有不同的物理意义。在生物学中，它不仅可用于描述肿瘤生长，也可描述一般组织生长[14,15]。图 5.1.8 表示不同时刻的波前波

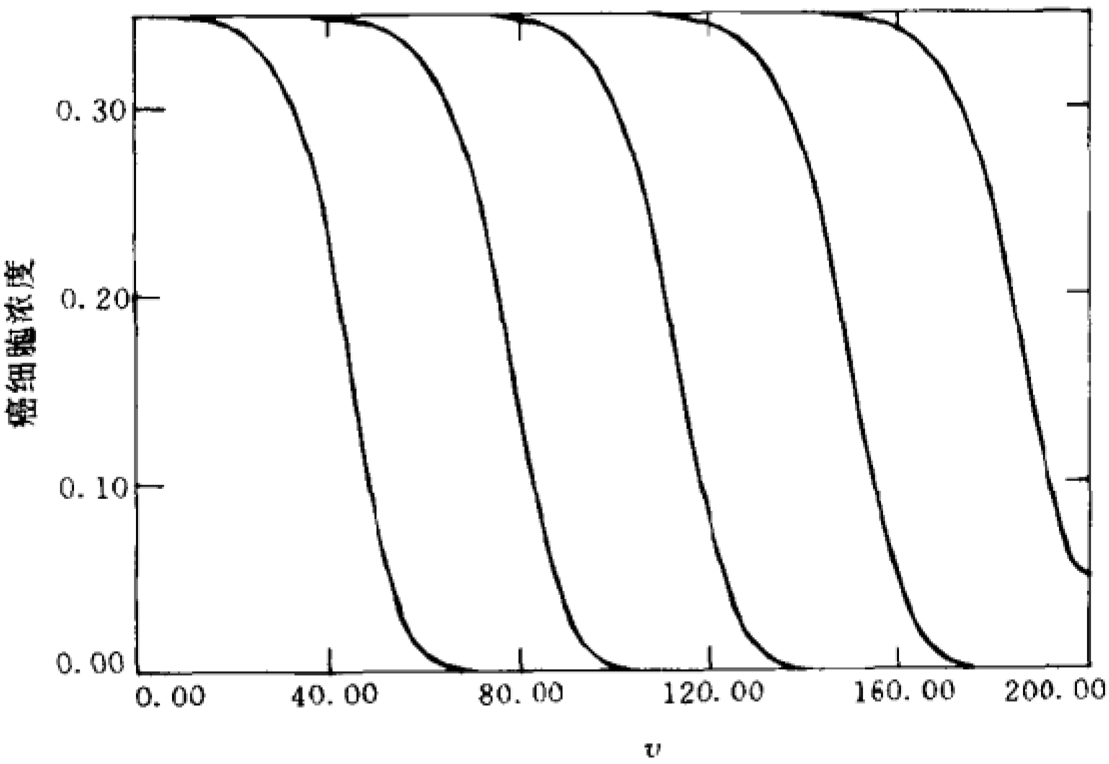


图 5.1.8 用波前表示癌的生长

诸曲线表示不同时刻的波前波形。曲线上面达到癌状态，下面则达到正常状态。曲线部分代表正常组织与癌组织的边界层。波以波速 v 传播意味边界层的移动即癌的生长。（引自文献[17]）

形。纵坐标与横坐标分别表示细胞或癌细胞的浓度和空间坐标。波前波形的弯曲部分表示正常组织或癌组织的边界，波前的传播代表边界的移动即组织的生长。波速具有重要的生物学意义，它反映组织生长的快慢。上述含四变量的非线性方程组有一种仅含两方程的形式，它的解更为丰富，不仅具有波前解还有极限环解[16]。最近的研究表明，只要对抗癌免疫作出更好的分析，对模型进行改进，完全有可能得到可以与实际肿瘤生长快慢相比较的波前波速，从而使理论贴近实际[17]。此外，还有人在 Lefever 等人工作的基础上考虑了更多的因素，用细胞自动机的方法研究在免疫监视下的肿瘤生长[18]，例如营养供应受到的限制，肿瘤生长时内部和外部密度的差别，部分细胞在细胞循环中进入静息状态等因素，能得到可与实验相比的结果。图 5.1.9(A)表示老鼠的 carcinoma KHT 的时间演化。光滑曲线 $V-t$ 为 Gompertz 曲线，来自于实验，V 表示体积。颤抖的曲线 N_0-t 代表理论结果，N_0 表示细胞自动机给出的肿瘤内的细胞总数。在理论与实验的比较中 V 和 N_0 都被归一化了。计算机实验还显示出癌的大小、总细胞数和其中包含的癌细胞数同时达到最大的稳定大小（图 5.1.9(B)）。

Lefever 曾研究参数涨落对肿瘤生长的影响。运用按动力学方程写出

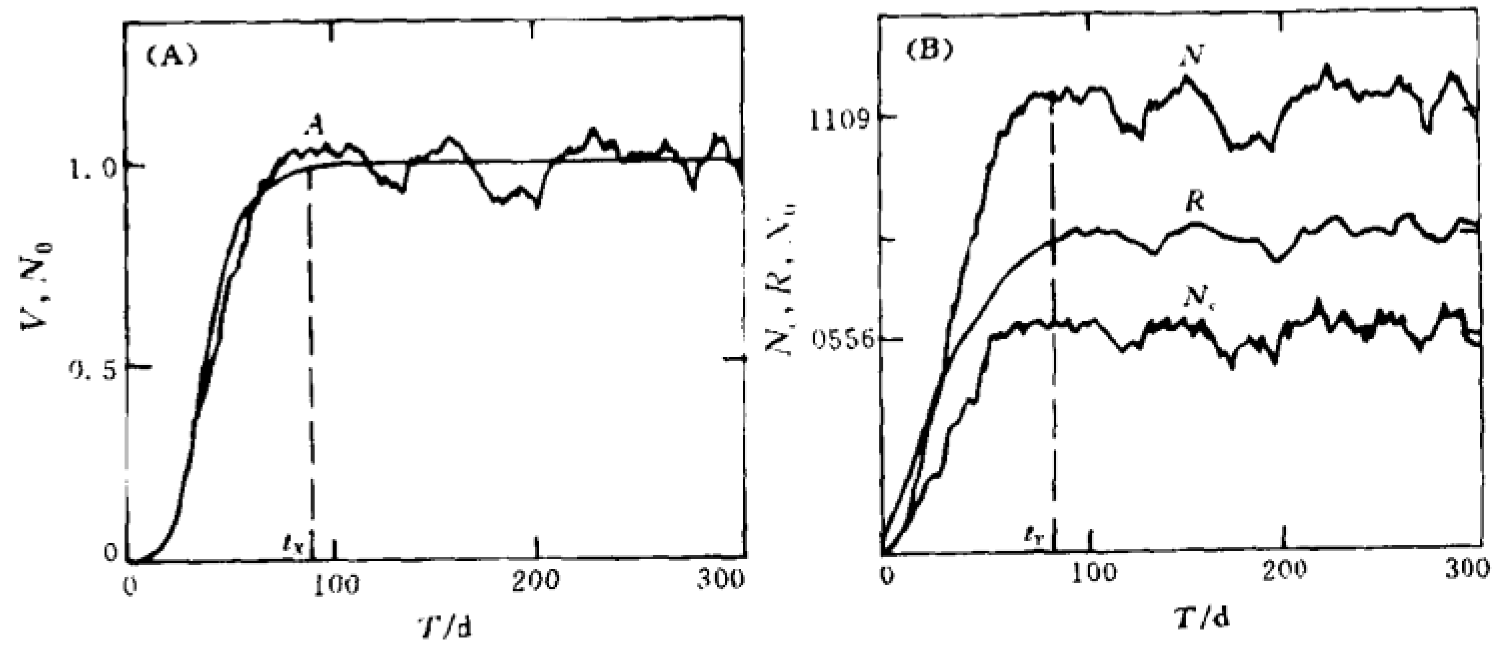


图 5.1.9　利用细胞自动机研究免疫监视下的肿瘤生长

(A) 关于鼠 carcinoma KHT 生长的理论结果与实验结果的比较；(B) 癌细胞数 N_C、癌组织的大小 R 和癌内各细胞总数 N_0 大体上同时达到各自的稳定值。

Ito 或 Stratanovich 意义下的随机微分方程和对应的 Fokker－Planck 方程并求定态解的方法，发现癌细胞增殖率的涨落不利于肿瘤的发展。运用细胞自动机讨论这个问题时，发现与上述一致的结果[19]。

以上讨论所采用的方法系考虑一复杂大系统，它包含若干子系统，根据诸子系统间的相互作用，经简化后用非线性数学的方法描述子系统间这种相互作用，研究各子系统或整个大系统的时空演化行为。例如研究考虑到独特型网络调节，从而得出诸种抗体数量随时间的演化。在某些情况下，上述方法也能揭露微观机制和宏观过程中的关系。例如肿瘤生长为宏观过程，而癌细胞的增殖、效应细胞对癌细胞的结合和杀伤等则属于微观过程。上述抗癌免疫监视的数学模型恰好能够给出各种微观过程，对癌组织整体的发展作出的贡献。非线性模型的方法有助于揭示同一层次不同组分间的协同作用以及不同层次的宏观与微观过程间的关系。这类工作进一步对生命现象和过程作出系统的理论性的解释，有助于实验工作的策划，并可能给出带预言性质的理论性的结果。

参考文献

[1]　林飞卿，余传霖，何球藻主编. 医学基础免疫学. 上海：上海医科大学出版社，1993

[2]　Bell G. *J Theor Biol*, 1970, **29**: 191

[3] Bell G. *J Theor Biol*, 1971, **33**: 339
[4] Fishman A M, Perelson A S. *J Theor Biol*, 1993, **160**: 311
[5] Kepler T. B, Perelson A S. *J Theor Biol*, 1993, **164**: 37
[6] Perelson A S. *SFI Studies in the Science of Complexity*, *Lect*. Vol. II, Ed. Erica Jan, Addison-Wesley, 1990.465
[7] Farmer J D, Packard N H, Perelson A S. *Physica 22D*, 1986,187
[8] De Boer R J, Kevrekidis I G, Perelson A S. *Chem Engng Sci*, 1990, **45**: 2375
[9] Zhang Libing, Du Chanying, Qi Anshen. *Bull Math Biol*, 1994, **56**(2): 323
[10] Varela F, *et al*. *Proc Natl Acad Sci USA*, 1991,**88**: 5917
[11] 杜婵英,孙祝,周丰群.北京师大学报(自然科学版),1996,**22**(1): 87
[12] Gray D, Leanderson T. *Expansion*, *Immunological Memory*. *Gray D(Ed.)*. Berlin: Springer-Verlag, 1990
[13] Guan Shuguang, Qi Anshen. *Bull Math Biol*, 1995, **57**(5): 713
[14] Lefever R, Erneux T. *Nonlinear Electrodynamics in Biological Systems*, W Ross and A Lawrence(Eds.). Plenum Publishing Corporation, 1984
[15] Qi Anshen. *Commun in Theor Phys*, 1988, **9**(2) 235
[16] Qi Anshen. *Bull Math Biol*, 1988, **50**(1): 1
[17] 周丰群,杜婵英,漆安慎.第十届全国非平衡统计物理学术论文集,载西北建筑工程学院学报.1996,(3):165
[18] Qi Anshen, Zheng Xiang, Du chanying, An Baosheng. *J Theor Biol*, 1993, **161**: 1
[19] 郑翔,谢冰蓉,杜婵英,漆安慎.北京师大学报(自然科学版),1997

(漆安慎)

§ 5.2 心脑可激媒质非线性电动力学

心脑可激媒质非线性电动力学是一个全新的研究领域。本节结合本课题组的研究内容介绍可激振子的耦合、可激媒质中的波传导、源场关系的逆问题和时间序列分析四个方面。这些内容既有长远的理论意义,又有广阔的应用前景,其中有许多十分诱人的课题,需要生物医学与数理科学跨学科的合作研究。在即将进入以生命科学为先导的 21 世纪的今天,心脑可激媒质非线性电动力学似乎进入全面深入的理论收割季节了。

5.2.1 理论描述

连续媒质电动力学所涉及的媒质是不可激发的,其电活动是线性的。可激媒质非线性电动力学与连续媒质电动力学之不同有两个基本点:一是媒质的可激性,二是电活动的非线性。所谓媒质的可激性是指动作电位

在媒质中可等幅度传导下去，但并不意味动作电位不消耗能量，而是传到哪里，哪里的媒质即可被激发，并通过主动输送过程，将其所消耗的能量给以补偿。所谓电活动的非线性具体表现在电流、电导随电压、时间的依从性，以及媒质对动作电位刺激反应的阈值性和不应性。

半个多世纪来，生物医学专家们在心脑可激媒质电活动方面，积累了大量的实验资料，建立了若干个适用于心肌和脑组织的数学模型，首先必须提到生物电学创始人，诺贝尔奖获得者 Hodgkin 和 Huxley(1952)[1]，他们在总结枪乌贼神经巨轴突膜动作电位曲线的实验数据基础上，引入膜通道开关因子，建立了描述神经膜动作电位的非线性微分方程。为了刻画膜电导对膜电位和时间的非线性，Hodgkin 和 Huxley 采用了常微分方程组的数值积分和电压箝位实验技术，结果表明数值积分重建动作电位的性态与实验描记的动作电位曲线符合得非常好，从而为神经生物电学或神经电生理学奠定了理论基础。Huxley(1959)[2]又将这一工作扩展到细胞外低 Ca^{2+} 溶液条件下，导致细胞膜自动除极和动作电位反复发生。这一工作提示许多人采纳 Hodgkin 和 Huxley 数学模型的框架，描述心肌可激媒质的电活动。

1962 年 Noble[3]将 Hodgkin 和 Huxley 方程应用于心肌 Purkinje 纤维动作电位研究，考虑舒张期电位的特点，引入 K_1 和 K_2 两个 K^+ 通道的电流 i_{K_1}(内向)和 i_{K_2}(外向)和 Cl^- 离子电流继用了 n,m,h 三个开关因子，并对 K^+ 电导和速率常数作了修正，研究了 Purkinje 纤维动作电位的重建问题。

1975 年，McAllister，Noble 和 Tsien[4]引入了 i_{Na}、i_{Si}(主要是 Ca^{2+})和 i_{qr}(Cl^-)三个电压-时间依从性内向电流和 i_{K_2}(舒张期)、i_{x_1}和 i_{x_2}(平台期)三个电压-时间依从性外向 K^+ 电流，及其相应的 6 个通道开关因子，建立了 MNT 方程，研究了心肌 Purkinje 纤维动作电位的重建问题，提出外向 K^+ 电导的减少对心脏起搏点电位起主要作用。这个论断后来得到 Noble(1983,1984)[5,6]，Difrancesco(1987)[7]等的重新解释。

1977 年，Beeler 和 Reuter[8]引用了四个离子电流，其中两个电压-时间依从性内向电流，即快 Na^+ 电流 i_{Na}和慢 Ca^{2+} 电流 i_S，另外还有两个外向 K^+ 电流，一个是时间独立的 i_{K_1}，另一个是电压-时间依从性的 i_{x_1}，引入了相应的 6 个通道开关因子，建立了 B－R 方程，研究了心室肌纤维动作

电位的重建问题，论述了心室肌纤维动作电位除极相、平台相和复极相的形成机理。

1980年，Kaoru Yanagihara，Akiniori Noma 和 Hiroshi Irisawa[9]，引用了四个电压-时间依从性电流：内向钠电流 i_{Na}。慢内向钙电流 i_S，外向钾电流 i_K和外向超极化活性电流 i_h，引用了相应的6个开关因子，建立了KAH方程，摸拟了窦房结(S－A node)细胞自发起搏的动作电位。结果指出，窦房结细胞动作电位的除极及舒张期自动除极主要由于 i_S内流，而平台相是由 i_S的失活和 i_K的激活造成的。

1985年，Difrancesco 和 Noble[10]研究了心肌 Purkinje 纤维舒张期自动除极相的机理，对 Noble 方程(1983，1984)给以修正，引用了相应的9个开关因子，建立了D－N方程。结果指出，心肌 Purkinje 纤维舒张期自动除极相的机制可能有：①外向 K^+电导自动减少；②内向 Na^+电导自动增大；③内向 Ca^{2+}电导自动增大。McAllister，Noble 和 Tsien(1975)提出①对自动除极相起主要作用；②对 Purkinje 纤维自动除极相的影响更大。

上述所有的数学模型都是对心脑可激媒质细胞膜的动作电位在离子通道开关因子控制下描述的，其共同形式为

$$
\begin{aligned}
&Cm\frac{\mathrm{d}V}{\mathrm{d}t}=\sum g_i(V,n_j)(V-E_i)\quad i=1,2,3\cdots,\\
&\frac{\mathrm{d}n_j}{\mathrm{d}t}=\alpha_{nj}(V)(l-n_j)-\beta_{nj}(V)n_j\quad j=1,2,3\cdots,
\end{aligned}
\tag{5.2.1}
$$

式中 V——膜电位，g_i——第 i 个离子通道的电导，n_j——第 j 个开关因子，E_i——第 i 个离子的平衡电位，α_{nj}、β_{nj}——开关因子速率常数。这是比较复杂的非线性常微分方程组，通常采用 Runge－Kutta 四阶近似数值积分法求解。

心脑可激媒质是由大量的细胞组成的，其动作电位的振荡在细胞膜上是靠局部环行电流的作用下传导的。在细胞之间也是靠局部环行电流的注入传递的。心脑可激媒质中动作电位波的尺寸比细胞的几何尺寸要大得多。如心室肌中的除极面的厚度，即动作电位波前与波后的间距约0.5mm，而心肌细胞的长度约100μm，直径约15μm。因此，完全可以将心脑媒质看成是连续媒质，引用“微观大宏观小”体积元，定义局域平均膜电位，胞内电导率，胞外电导率和空间场量等，建立心脑可激媒质动作电位

波的运动方程：

$$\begin{aligned}\frac{\partial v}{\partial t} &= D\nabla^2 V + I(V,\boldsymbol{n}),\\ \frac{\partial n_j}{\partial t} &= \delta\nabla^2 n_j + f(V,n_j),j=1,2,\cdots,\end{aligned}\tag{5.2.2}$$

式中，V 是膜电位，$n(n_1,n_2,\cdots)$是膜通道开关因子，等号后第一项是扩散项，第二项 $I(V,\boldsymbol{n})$是膜电流非线性项，$f(V,n_j)$是开关因子的变化速率，D,δ 是扩散系数，对心脑媒质 $\delta=0$。这心脑可激媒质动作电位非线性偏微分方程的形式与化学等可激媒质的反应-扩散方程相同，全面求解异常困难。化学可激媒质对此有些简化求解的方法可以借鉴。

5.2.2 研究内容

心脑可激媒质非线性电动力学的研究内容，在源的方面有：①可激振子的耦合，②可激媒质中的波传导；在场的方面有：③源场关系的逆问题，④时间序列分析。

1. 可激振子的耦合

由多细胞组成的某些集团，其电活动往往是同步化了的，同单个细胞具有相类似的可激性，如心脏的窦房结、房室结及脑组织的分区单元等。因此可将这些多细胞集团模拟成一个可激振子。通过对这些可激振子相耦合的探讨，可方便地研究多细胞集团电活动间的相互作用。心脏中常分有本位起搏点，即窦房结；异位起搏点，如房室结区和心室肌某处 Purkinje 纤维的兴奋点等。心脏异位起搏点与本位起搏点的相互作用是可激振子耦合的重要研究内容。

多数生物可激振子电活动极限环具有较强的吸引域，许多生物振荡的数学模型具有强吸引域的稳定极限环。在大多数情况下，给起搏点振荡以单脉冲扰动，将对进行中的节律产生相位重置。这重置的大小不但取决于刺激强度，而且还依赖于刺激在周期中的相位。这新相与旧相的函数关系叫做相位响应曲线(PRC)。相位响应曲线的分支数可以是1(奇重置)或0(偶重置)的连续函数，也可以是不连续函数。一般来说，在低刺激强度下给出奇重置的PRC，在高刺激强度下给出偶重置的PRC。

当起搏点的固有节律受到周期性扰动时，随着扰动频率和强度的改变，在扰动与起搏点之间会建立起不同的耦合模式。在某些情况下，起搏

点振荡的节律被锁定在周期扰动的某一固定的相位处，结果对每 N 个扰动周期，存在 M 个起搏点振荡节律周期，即 $N:M$ 锁相(phase locking)。最常观测到的稳定锁相区相当于低阶锁相比，即 2∶1，3∶2，1∶1，2∶3，1∶2 等，高阶锁相比也可以观测到，但它们在频率-强度参量空间中仅占据较少的区域，因而容易被忽视或被噪声所掩盖。研究频率-强度参量空间上稳定锁相区之间的区域，往往能够找到产生不规则动力学特性的扰动参量。对周期性扰动非线性振荡器的研究有着丰富多采的历史，至今该领域结合生命科学的研究非常活跃。在心脏方面代表性的实验和理论研究工作有 Guevara，Glass 和 Shrier (1981)[11]；Glass 等 (1983，1984)[12,13]；Guevara(1984)[14]；Glass，Guevara 和 Shrier(1987)[15]。

在耦合比较弱的情况下，设时间 T_i 起搏点振荡的有效相位为 Φ_i，经过一个周期性扰动的周期 T_0 后，T_{i+1} 时间起搏点振荡的有效相位为 Φ_{i+1}，前后相位关系可用如下差分方程表示：

$$\Phi_{i+1}=\Phi_i+f(\Phi_i,b)+\Phi_0 \quad (\bmod 1),$$

在此式中函数 $f(\Phi_i,b)$ 通常是依赖于耦合强度 b 的非线性函数，当 $b=0$ 时，$f(\Phi_i,0)=0$，Φ_0 是与周期 T_0 相当的相位。只要已知非线性函数 f，就能够计算出所有将来时刻的起搏点振荡的动力学特性。此方程的迭代过程相当于把单位圆上的一个点映射到第二个点，这种函数叫做圆映射。为了研究扰动的效应，重要的是了解当圆映射中参量改变时，在参量空间中受扰起搏点振荡动力学特性发生的变化(即分岔)。以耦合强度 b 为纵坐标，以周期性扰动周期相位 Φ_0 为横坐标，在此平面参量空间(b,Φ_0)上，存在一些 $N:M$(扰动周期数与起搏点振荡周期数之比)的稳定锁相区域叫作 Arnold 舌头或喇叭。

受周期性扰动的起搏点振荡节律，其相位可锁定于扰动周期的相位上，也可最终变为非周期性的。在出现非周期性节律情况下，扰动强度较低时观察到拟周期节律或同步定律；扰动强度较高时，非周期性动态有时会与混沌相联系。锁相区是扰动频率和强度的函数，其排布方式表现出许多显著特征。在扰动强度较低时可观察到 Arnold 舌头结构。此结构中存在和比定则：若在一频率处有 $N:M$ 锁相，在另一频率处有 $N^*:M^*$ 锁相，则在两套之间的一个中间频率上可以找到$(N+N^*):(M+M^*)$锁相。随着刺激强度的增大，Arnold 舌头结构被破坏，并产生复杂的拓扑结

构。

临床电生理学认为锁相现象十分重要，在许多电活动中都能观测到锁相现象。Moe 等（1977）对窦性心律与异位起搏点之间的稳定锁相区进行了计算[16]。张广才等对周期性扰动下可激系统动力学及异常心律给出符号动力学研究。由于有些异位起搏落在心室的不应期内，不是所有的异位起搏都能观测得到。临床中隐匿性二联律和三联律就是这个道理。房室传导电活动可用窦房结、房室结区和心室三个非线性振荡器来建模，通过减少房室结区与心室间的耦合强度，有可能得到许多不同的稳定锁相节律，这些节律在临床中定性地对应于一类叫作房室传导阻滞。

2. 可激媒质中的波传导

对非线性偏微分方程全面求解异常困难，研究其行波解，建立随波运动的标架，变为常微分方程或带小摄动的常微分方程求解却容易得多。

心脏是一个复杂的三维构造。但其中有兴奋传导系统，即动作电位波是按如下次序传导的：窦房结（S－A node）—心房（A）—房室结（A－Vnode）—His（束）（H）—Purkinje 纤维（P）—心室（V）。动作电位波沿此传导系统可视为一维组织中的传播。心脏电生理学中还有一个重要机理——折返传导，即动作电位围绕一个孔或障碍物（胞膜间隙、房室旁路，动脉瘤或梗死灶等）呈环形波在心肌组织中反复传播。动作电位波沿此折返回路传导亦可视为在一维组织中的传播。

（1）稳定的孤波解

在无边界系统中的局部激发，即

$$\begin{aligned} V(x \to +\infty) &= V(x \to -\infty) = V_0, \\ \boldsymbol{n}(x \to +\infty) &= \boldsymbol{n}(x \to -\infty) = \boldsymbol{n}_0 \end{aligned} \tag{5.2.3}$$

条件下，系统将自动演化为稳定的行波，从某种意义上讲，行波解具有一定的普遍性。实际上是解传播速度的本征方程，系统的可激性可由传播速度 $c(V_0, \boldsymbol{n}_0)$ 定义，其中 $(V_0, \boldsymbol{n}_0)$ 就是方程的不动点，$(V_0, \boldsymbol{n}_0)$ 决定系统的可激性质。

（2）平面波波列

对无外加扰动的波列，可借助于孤波解的微扰得到其演化性质，即

$$\Phi = \sum \Phi_i[x - ct - \zeta_i(t)] + \varepsilon R \tag{5.2.4}$$

式中 Φ_i 代表每个波，$\zeta_i(t)$代表每列波相对传播的微运动。这些平面波列将自发调节其间隔以达到稳定间隔。对孤立波尾是振荡衰减的情况，还会引发波之间相对振荡的离合运动。

对外加扰动的强迫型的波列，其周期 T 受外源控制，在一定的 T 区间内，波列中每一个波与孤立波形相差不大，唯一区别就是恢复时间 T 对波列传播速度的调制。通过规定周期边界条件，解关于传播速度 c 的本征方程，可求得 $c=c(T)$的关系。随着扰动频率的增长，波的传播速度将减慢。传播速度与频率相关的这种现象，就是物理学上的色散关系。

一维波传导的研究可增进临床对房室传导阻滞，二度阻滞中的 Wenckebach 型和 Mobitz 型节律以及室上性心动过速高频的 QRS 窦性波与复合波的交替节律等动力学特性的认识。

早在 1924 年，Garrey 在研究海龟心脏时，就对没有障碍物的心脏组织中存在环行波做过描述。1946 年，Wiener 和 Rosenblueth 对折返和室颤首次进行了基础理论的研究[17]。1948 年，Selfridge 亦证明 Wiener 和 Rosenblueth 模型在没有障碍物的情况下可能维持稳定的折返运动[18]。

Winfree(1980，1987)对有关生理学中二维波传播方面的主要工作做了完整的总结[19,20]。其中在 Belousov - Zhabotinsky 化学可激媒质反应产生螺线波的讨论中明确指出，即使在没有障碍物的情况下，二维组织中也会出现稳定环行运动的螺线波，并经以后工作严格证实，这对在没有障碍物(梗死灶等)情况下出现的折返性心律失常提供了理论解释的依据。Farley 和 Clark(1961)首次在神经网络模型中观察到了螺线波[22]。Moe 等(1964)[23]及 Smith 和 Cohen(1984)[24]在心肌元胞自动机模型的数值仿真中观察到有螺线波。

(3) 曲面波的传播

对靶波和旋波传播的理论研究，可建立随波前一起运动的标架，将其化为局部的一般传导问题，同样可以得到传播速度。不过，这里的传播速度要比一维情况稍复杂一些，传播速度 c 与波面的位形 k(曲率)有关，即 $c=c(k)$。对平面上的曲面波来讲，当 k 较小波前不是太挤的情况下，$c=c_0-Dk(k\ll c_0/D)$。

对断裂的波前，波前和波后在此处连接，根据此处的曲率情况，确定尺度变换关系。用平面断波的演化的微扰，可得到波端点的曲率与纵向生

长速度之关系。对于 k 较小时，有 $c_1=c_{10}-\gamma k_0$，k_0 是波端点的曲率，即 $k(l\to0)=k_0$。

这里有三个理论问题需要处理：

① 曲面波线演化的描述问题

上述公式分别给出了波传播速度 c 与 n_0、T 和 k 的关系。这样，空间任一曲面波波前的演化情况均可根据每点处波前的传播速度 c 加以描述。在弱可激情况下，l/k 远大于波前区域，平面上曲面波的波前则可用 $k(l)$来描述，其中 l 为曲面波线的弧长，根据 $c=c_0-Dk$，可得到[24]：

$$\frac{\partial k}{\partial t}+\left(\int_0^l kc_0\mathrm{d}\xi+c_1\right)\frac{\partial k}{\partial l}=-k^2c_0+D\frac{\partial^2 k}{\partial l^2},\qquad(5.2.5)$$

可用此式讨论波前的演化，也可求解波前稳定运动的性质，如对稳定的旋波 $\frac{\partial k}{\partial t}=0$，利用 $c_1=0$，波端的生长，可得到自由旋波核的尺寸 R。再利用 $k(l\to\infty)=0$，$k(l\to0)=k_0$，可得到旋波转速关系。当然，也可通过它研究非均介质 $c_1(\vec{r})$，$c_0(\vec{r})$或扰动下的媒质 $c_1(t)$，$c_0(t)$的演化性质。如果考虑恢复时间 T 的作用，则可得 $c_0(T)$。当然，一般 T 较大时，$c_0(T)=c_0$。但当 T 小时，如在波端处，转速大会对 c_1 和 c_0 产生影响。一般情况下，可使旋波产生曲屈运动。

② 旋波核的处理问题

为了研究旋波结构的稳定性，对可激媒质旋波的渐近解要分成外区与“核”区两个区域分别进行处理。当转速与 k_0 较大时，上面提的一套方法不再适用，对此要用数值方法进行研究或处理。

③ 旋波破碎问题

曲面波传播分析表明，当波线曲率 k 很大，恢复时间 T 很小时，上次传来的波媒质来不及恢复，则可能引起下次传来的波形发生变化。特别是当波线长参数 l 减少时，可使波发生分裂，产生大小不同的波段。旋波在一定条件下发生分裂，产生更多的旋波，这叫旋波破碎(fragmenting of spiral wave)。旋波破碎表现在心脏中是多个中心的除极，这可能是心律失常的机制；旋波多次破碎形成自旋玻璃态，也许是心室纤颤的基础。心脏的各向异性及其有限尺寸是否对室颤的产生有根本影响，这一点尚未研究清楚。

实际上，心脑可激媒质的兴奋波多是在三维空间中传播的。但是，对

三维组织（心室和大脑皮层结构）中的兴奋波进行观测是相当困难的，目前得到的主要结果大多是理论性的，主要进展是由 Winfree 和 Strogatz (1983,1984)取得的，用计算机作出了三维卷曲波[25~29]。Winfree 于 1987 年又对此做了详细的总结和解释[19,20]。要评论这些理论的实际意义为时尚早。然而，在化学可激媒质的 BZ 反应中确实观察到了这理论的预期结果，在心肌旋波理论和观察方面，有迹象表明亦会如此。

3. *源场关系的逆问题*

上面讲的心脑可激振子的耦合和可激媒质中的波传导都是在源的方面研究问题。源场关系的逆问题和时间序列分析是在场的方面讨论问题。

心电导联有常规导联、Frenk 正交导联和多道标测导联；脑电有常规 16 导联等，都是描记场的电位或电压变化的。

由心脑可激媒质中电活动（源）讨论心脑电图（场），这是正问题，便于论述心脑电图的形成原理；而由心脑电图（场）研究心脑电活动（源），这是逆问题，便于探查心脑组织是否正常。Humphrey(1968)[30]，Freeman 和 Nicholson(1975)[31]，及 Mitzdorf 和 Singer(1978)[32]研究 EEG，描记头皮电流源与穴的强度和分布。Katznelson (1980)[33]，Hjorth (1979)[34]，Chapman 和 Evans(1976)[35]，及 Freeman(1980)[36]研究 EEG，定量探测颅内深部脑电流源偶极矩的大小和方位。我们课题组从事心性猝死 SCD 的研究，定量地计测心脑组织缺血，损伤面积和坏死体积的大小和方位[37]，上述这些都是源场关系中逆问题的基础或应用基础研究。

4. *时间序列分析*

心脑媒质系统是复杂系统。复杂系统信号的时间序列分析，重构相空间具有其很独特的意义。因为往往在无穷多自由度系统中，其运动模式却是有限维的，这在心脑波传导图象的形成中很常见。根据 Taken 的重建原则对系统的复杂性进行描述，对于处理脑电，心电及其他生物电信号性质都是很有用的。

由于神经元的大数性及细胞间突触联接的复杂性（化学递质、电场等），对脑媒质作出完整的描述可以说是不可能的。对人脑动作电位和皮层脑电信号的时间序列分析表明，在某些简单情况下，大脑不是工作在周期振荡上，而是工作在一个维数不高的混沌吸引子上。这可能是因为生物

要进行敏捷的反应,要尽量防止其锁模运动;而具周期运动的系统往往出现多种模式的锁模形式。对动物大脑切片的测试表明,在重构的空间中,的确有很多双曲型的不动点,接近这些不动点的点具有收缩流形和放大流形运动的性质,这可以同随机运动明显区分开来。通过对这些接近不动点的点的信号加入小摄动,可成功地将其控制在单周期或小周期上。

心电的处理也具有特别意义。心电 $R-R$ 间期对正常人来说是有漂移的,$I_n=RR_{n+1}-RR_n$ 作为信号的时间序列分析表明,I_n 具有长程的关联。而对有心脏病的,这种关联弱,对新生儿来说,也是缺乏此种关联的。这表明这种关联对应的是神经慢调节,这种慢调节对维持心脏节律是很关键的。对心律失常的心电图研究,其短时间呈现的周期行为,可以指导临床对心脏起搏点性质作出分析。由本位起搏点(窦房结)和异位起搏点共同作用的心室搏动可能是周期性的,但不一定是常节律的。

5.2.3 应用前景

心脏可激媒质非线性电动力学研究有个突出特点:既有长远的理论研究目标,又有广阔的应用前景。我们结合生物医学研究主要集中于:①心性猝死 SCD(I)——心肌梗塞缺血与损伤面积和坏死体积计测数学模型、动物实验和临床测试的基础研究[37,38];②心性猝死 SCD(II)——心肌可激媒质非线性电动力学及其在消除心室纤颤的理论基础研究;③心性猝死 SCD(III)——控制混沌与心室纤颤监测、预报和诊治的基础研究;④脑电混沌动力学的研究[39]。

心性猝死在心电图上表现:正常→室早→室速→室颤→猝死,其动力学过程是:正常节律→病态周期振荡→混沌动态→垂死周期振荡→猝死不动点。SCD 在美国每年要夺去 40 万,在我国每年要夺去百万余人的生命。10 年来国外有几个研究中心从事 SCD 的监测预报和防治的研究工作。Lown(1979)论述了关于 SCD 的各种流行的观点[40];Nikolic 等(1982)[41]和 Hohnloser 等(1984)[42]讨论了室早对引发室颤导致 SCD(R on T 现象)的作用;Winfree(1983,1985,1987)提出有关室速和室颤起源的假说[43~45],落在不应期上的单个刺激可以引发螺线波,如果刺激所产生的一个波与正在传播的另一个波的不应区尾部相交,那么后者的端点就可能会卷起来产生两个相反方向旋转的螺线波。因此,给自发振荡的均匀组织加单脉冲有可能导致分级相位重置,最终引发螺线波。Ideker 和

Shibata(1986)采用多导检测技术检验了 Winfree 的假设[46]，发现实验结果与理论预测相一致。

室颤是否是混沌的，一直是 Battersby(1965)[47]、Smith 和 Cohen(1984)[48]和 Goldberger 等(1986)[49]争论的一个焦点。由于目前对室颤和混沌都没有公认的可操作性定义，看来这个问题还要争论下去。

Mackey 和 Glass(1977)[50]首创"动态病"(dynamical diseases)名词，之后经常讨论这个问题。动态病是以非寻常的复杂动态行为为特征的疾病。分析引起动态病的机理，不可避免地要对观察到的动力学特性作理论分析。研究这些问题的途径是建立疾病的理论模型和医学模型。研究动态病可为人类预防、诊断和治疗提供新概念、新原理和新方法，这是数理专业与医学专业的专家们协作研究的重要领域[51~54]。

参 考 文 献

[1] Hodgkin A L, Huxley A F. *J. Physiol. (Lond)*, 1952, **117**: 500

[2] Huxley A F. *Ann. N Y Acad Sci*, 1959, **81**: 221

[3] Noble D. *J Physiol*, 1962, **160**: 317

[4] McAllister R E, Noble D, Tsien R W. *J Physiol*, 1975, **251**: 1

[5] Noble D. *Symp Soc Exp Biol*, 1983, **37**: 1

[6] Noble D. *J Physiol (Lond)*, 1984, **353**: 1

[7] Difrancesco D. *J Physiol (lond)*, 1984, **348**: 341

[8] Beeler G W, Reuter H. *J Physiol*, 1977, **268**: 177

[9] Kaoru Yanagihara, Akiniori Noma, Hiroshi Irisawa, 1980, *Japanese J Physiol*, 1980, **30**: 841

[10] Difrancesco D, Noble D. *Phil Trans R Soc (Lond) B*, 1985, **307**: 353

[11] Guevara M R, Glass L, Shrier A. *Science*, 1987, **214**: 1350

[12] Glass L, Guevara M R, Shrier A, Perez R. *Physica 7D*, 1983: 89

[13] Glass L, Guevara M R, Belair J, Shrier A. *Phys Rev*, 1984, **29**: 1348

[14] Guevara M R. *Chaotic Cardiac Dynamics*. Ph. D dissertation, McGill University, 1984

[15] Glass L, Guevara M R, Shrier A. *Ann N Y Acad Sci*, 1987, **504**: 168

[16] Moe G K, Jalife J, Mueller W J, Moe B. *Circulation*, 1977 **56**: 968

[17] Wiener N, Roseenblueh A. *Arch Ins Cardiol Mex*, 1946, **16**: 205

[18] Selfridgee O. *Arch Ins Cardiol Mex*, 1948, **18**: 177

[19] Winfree A T. *The Geometry of Biological Time*. New York: Springer-Verlag, 1980

[20] Winfree A T. *The Time of Biological Clock*. New York, 1987

[21] Farley B G, Clark W A. Activaty in networks of neuronlike elements, Information Theory, 4th London Symposium. ed. C Cherry. London: Butterworths, 1961

[22] Moe G K, Rheinboldt W C, Abildskov J A. *Am Heart J*, 1964, **67** : 200

[23] Smith J M, Cohen R J. *Proc Nat Acad Sci USA*, 1984, **81** : 233

[24] Mikhailov A S, Davydov V A, Zykov V S. *Physica D70*, 1994 : 1

[25] Winfree A T, Strogatz S H. *Physica 8D*, 1983a : 35

[26] Winfree A T, Strogatz S H. *Physica 9D*, 1983b : 65

[27] Winfree A T, Strogatz S H. *Physica 9D*, 1983c : 333

[28] Winfree A T, Strogatz S H. *Physica 13D*, 1984a : 221

[29] Winfree A T, Strogatz S H. *Nature* , 1984b, **331** : 611

[30] Humphrey D R. *EEG clin. Neuro - physiol*, 1968, **24** : 116

[31] Freeman J A, Nicholson C. *J Neuro - physiol*, 1975, **38** : 369

[32] Mitzdorf U, Singer W. *Exp Brain Res*, 1978, **33** : 371

[33] Katznelson R D. Bostom: American EEG Society Meeting, 1980

[34] Hjorth B. *Pharmakopsychiat*, **12** : 11

[35] Chapman L F, Evans J W. *EEG Clin Neuro - phsiol*, 1976, **40** : 324

[36] Freeman W J. *IEEE Trans Biomed. Eng*, 1980, **27**(8) : 421

[37] 曲直.心肌梗塞面积计测数学模型,生物物理学报,1991,**7** : **3**

[38] Qu Zhi, *et al*. *Mathematical Biology*, *Proceedings of the Iternational Conference on Biomathematics*. Xian Jiaotong University press, 1988. 147

[39] Dong Wei, Qu Zhi. *Study of Chaotic Dynamics of Brain Activity under an Anestthetic*, *Third China - Japan Bilateral Symposum on Biophysics*. China Xi'an, 1991

[40] Lown B. *Am J Cardiol*, 1979, **43** : 313

[41] Nikolic G, Bishop R L, Singh J B. *Circulation*, 1982, **6** : 218

[42] Hohnloser H, Weiss M, Zeiher A, Wollschlarger H, Hust, M H, Just H. *Clin. Cardiol*, 1984, **7** : 517

[43] Winfree A T. Sudden Cardiac Death: A problem in topology. *Sci Am*, 1983b, **248**(5) : 144

[44] Winfree A T. Orginizing centers for chemical waves in two and three dimensions. In : Oscillations and Traveling Waves in Chemical Systems, ed. Field R J and Burger M, 1985

[45] Winfree A T. When Time Breaks Down: The Three Dimensional Dynamics of Electrochemical Waves and Cardiac Arrhythmias. Princeton : Princeton University, 1987

[46] Ideker R E, Shibata N. Experimental evaluation of predictions about ventricular fibrillation based upon phase resetting. (Abstract), Vencouver, Int. Union Physiol. Sciences, 1986

[47] Battersby E J. *Circ Res*, 1965, **17** : 296

[48] Smith J M, Cohen R J. *Proc Nat Acad Sci*, USA, 1984, **81** : 233

[49] Goldberger A L, Goldberger E. Clinical Electrocardiography: A Simplified Approch. ed. St. Louis, Mosby C V

[50] Mackey M C, Glass L. *Science*, 1977, **197** : 287

[51] Glass L, Mackey M C. From Clocks to Chaos. The Rhythm of Life. Princeton University Press, 1988

[52] Holden A V, *et al*. Nonlinear Wave Processes in Excitable Media. New Plenum Press, York

[53] Zykov V S. Modelling of Wave Processes in Excitable Media. Hanchester Unversity Press

[54] Jalife J. Mathematical Approaches to Cardiac Arrhythmias, New York Academy of Sciences, 1989

(曲 直 张广才)

§5.3 分子马达——噪声驱动定向运动的一个模型

生物体的运动诸如肌肉收缩，细胞活动和细胞内的输运过程等，追踪到分子水平都是来自具有小马达功能的分子马达在起作用的结果，这些马达是蛋白质分子，因而也称为马达蛋白。它们能够高效率地将贮藏在三磷酸腺苷(ATP)分子中的化学能转换为机械能，产生协调的定向运动。有关这一能量转换的分子机制一直是分子生物学中最为引人注目的问题之一。

早期有关这方面的研究多在肌肉收缩及体内的一些过程，由于测量是间接的且又涉及多分子参与，因而不同测量的结果相差非常大，无法搞清问题。近年来实验技术突飞猛进，已有可能在实验室里直接观测到单个分子马达的运动，从而打开了一个新时期。

到目前为止已有近百种马达蛋白被确定，各种各样的马达蛋白大都是沿着相对应的蛋白丝作定向运动。这些蛋白丝起着轨道作用。例如肌球蛋白是沿着肌动蛋白丝运动，动蛋白是在微管上运动。这是迄今研究最多的两个系统，测定表明这些分子马达进行的是单方向与跳跃式的运动。

以前的理论模型大都建立在 Huxley 的“横桥模型”(crossbridge model)[1]基础上，如转动横桥模型。但缺乏有关肌球蛋白头部转动构象变化的实验证据，而其他马达蛋白，如动蛋白始终保持着与微管接触，更不容易设想有较大的构象变化。为了摆脱这一困境，近来又有人提出新的观点：马达蛋白沿蛋白丝通过单向扩散从上一单元到达下一单元。近来从动蛋白沿微管运动的实验中观测到，每一跳跃是 8nm[2]，恰好与微管的每一单元节长度相吻合。由此可见，这一观点有一定合理性。但是我们知道根据热力学第二定律，白噪声是不能用以产生定向运动和做功的。最近的一系列研究阐明了在非对称位势中有色噪声是可以驱动定向运动

的[3~7]。在这类模型中，假定有色噪声源来自 ATP，其中 Astumian 等[5]还考虑到由于 ATP 是带电的，它与 Brownian 粒子(马达蛋白)的结合会导致位垒高度的涨落从而引发定向流。因此他们研究了一个 Brownian 粒子(在过阻尼介质中)在一维不对称周期位阱(如锯齿型)和有色噪声或涨落力和涨落位垒作用下的运动情况。

进一步假定来自 ATP 的噪声不仅是有色的，且与其位置有关[8]。我们着重分析动蛋白与微管系统。实验证明[9]ATP 的附着及其水解均发生在动蛋白与微管脱离之前，因此可以合理地认为噪声与动蛋白和微管的结合点的位置有关。因而我们所研究的数学模型不再是加性噪声问题，而是乘性噪声的问题。

这里首先讨论一个较为一般的情况，使之不仅可以描述定向运动，且可同时描述跳跃性运动，以符合实验结果[2]。

考虑一个处在过阻尼环境中的 Brownian 粒子在一个不对称的一维周期位势(包括随时间变化)和乘性有色噪声驱动下的运动过程，我们可以写出如下的 Langevin 方程：

$$\dot{x}=-\frac{\partial}{\partial x}V(x,t)+\xi(t)+g(x,t)\varepsilon(t)。\tag{5.3.1}$$

这里 $V(x,t)$ 为一个包括随时间涨落变化的一维不对称的周期位势，位势的极小点代表分子马达与微丝(管)的结合点，周期长度 L 为微丝(管)的单元节长度。$\xi(t)$ 是白噪声，描述环境介质对 Brownian 粒子的作用，

$$\langle\xi(t)\xi(s)\rangle=2kT\delta(t-s),\tag{5.3.2}$$

而 $g(x,t)\varepsilon(t)$ 为乘性色噪声，来源于 ATP 的水解，且与作用点以及作用时间有关，由 $g(x,t)$ 来表示。

为了便于进行物理分析，我们首先研究下面几种特殊情况。

1. 仅考虑乘性噪声的作用[8]

$$\dot{x}=-\frac{\partial}{\partial x}V(x)+g(x)\varepsilon(t),\tag{5.3.3}$$

这里 $V(x)$ 与 $g(x)$ 均假定为具有相同周期长度的周期函数，这意味着 ATP 水解所处的空间位置与结合点是相同的，即化学过程快于力学过程。$V(x)$ 与 $g(x)$ 的形状见图 5.3.1。假定噪声是关联时间为 τ 的高斯

指数型色噪声，即

$$\langle \varepsilon(t) \rangle = 0, \tag{5.3.4}$$

$$\langle \varepsilon(t)\varepsilon(s) \rangle = \frac{D}{\tau}\exp\left(-\frac{|t-s|}{\tau}\right)。 \tag{5.3.5}$$

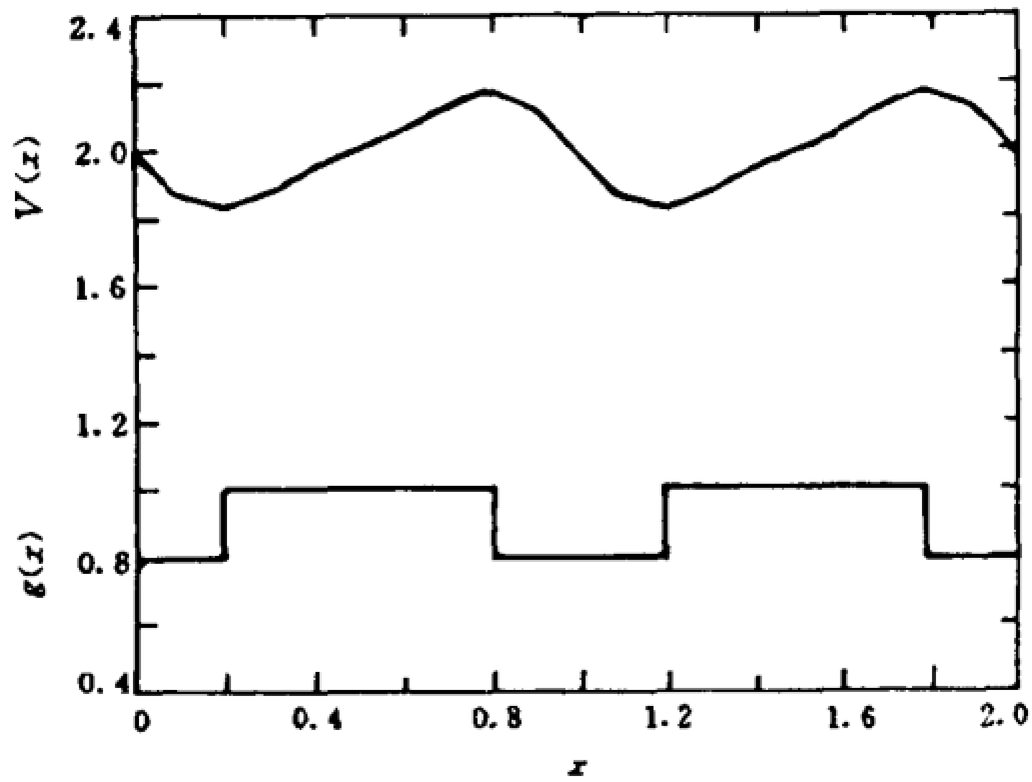


图 5.3.1 周期位势 $V(x)$及乘性噪声因子 $g(x)$的示意图

即使是这么简单的一维色噪声随机系统，由于是非马尔可夫过程，在一维空间不再遵循 Fokker - Planck 方程，只能在各种近似条件下导出各类有效的 Fokker - Planck 方程，或在二维空间，由二维马尔可夫过程来替代，写成二维的 Fokker - Planck 方程。

Millonas 等[4]对于加性色噪声在弱噪声近似（小 D 近似）下得到一个定态流的解析公式，能很清楚地说明问题。在相同近似条件下，我们推广到乘性噪声情况，仍然得到了一个定态流的解析表达式。

显然，当噪声强度 D 很小时，布朗粒子（马达分子）大部分时间都是处在位势的极小点 x_0 附近作小振幅涨落运动，但它偶然也会从那儿“跳”过其右边或左边位势的极大点 x_+或 x_-，其单位时间的概率分别为 W_+与 W_-，从而定态流为 $S=W_+-W_-$[4]

对于乘性噪声，只要 $g(x)\neq 0$，我们总可以做坐标变换，如令 $y=F(x)=\int_{x_0}^{x}\frac{\mathrm{d}x'}{g(x')}$，使原来对 x 坐标而言是乘性噪声方程，变到对 y 坐

标而言是加性噪声方程。即

$$\dot{y}=\frac{\dot{x}}{g(x)}=-\frac{V'(x)}{g(x)}+\varepsilon(t)=-U'(y)+\varepsilon(t), \tag{5.3.6}$$

这里

$$U'(y)=\frac{V'(x)}{g(x)}\bigg|_{x=F^{-1}(y)}, \tag{5.3.7}$$

而

$$U(y)-U(y_0)=\int_{x_0}^{x}\frac{V'(x')}{g^2(x')}\mathrm{d}x'\bigg|_{x=F^{-1}(y)}, \tag{5.3.8}$$

$$U''(y)=g(x)\left(\frac{V'(x)}{g(x)}\right)'\bigg|_{x=F^{-1}(y)}。 \tag{5.3.9}$$

由此可见$U(y)$与$V(x)$是一一对应的。由于在y坐标系中，方程已变为加性色噪声问题，我们可以直接应用 Millionas 等[4]的方法写出在y坐标系中的定态流，然后再变回到x坐标系中去。在小D近似中，$W_\pm$与D的关系可写为：

$$W_\pm=\text{const}\times\exp(-R_\pm/D), \tag{5.3.10}$$

其中$R_\pm$是从位势极小点x_0逃逸到位势峰点x_+或x_-的活化能，可由变分法得到[10]，$R_\pm=2F(0)[U(y_\pm)-U(y_0)]+r_\pm F''(0)$，这里

$$r_\pm=\int_{y_0}^{y_\pm}\mathrm{d}yU'(y)\,U''^{2}(y)\geqslant 0, \tag{5.3.11}$$

$$F(w)=D/\Phi(w),\quad F''(w)=\frac{\mathrm{d}^2F(w)}{\mathrm{d}w^2}, \tag{5.3.12}$$

$$\Phi(w)=\int_{-\infty}^{\infty}\mathrm{d}t\exp(\mathrm{i}wt)\Phi(t), \tag{5.3.13}$$

$$\Phi(t)=\langle\varepsilon(t)\varepsilon(0)\rangle。 \tag{5.3.14}$$

由于跃迁率具有坐标变换不变性[11]，所以回到原始的x坐标系，我们得到定态流为：

$$S=\frac{\sqrt{V''(x_0)V''(x_+)}}{2\pi}\left\{\exp\left[-\frac{2F(0)}{D}\Delta V(x_+)-\frac{r_+F''(0)}{D}\right]\right.$$
$$\left.-\exp\left[-\frac{2F(0)}{D}\Delta V(x_-)-\frac{r_-F''(0)}{D}\right]\right\}, \tag{5.3.15}$$

这里

$$\Delta V(x_{\pm})=\int_{x_0}^{x_{\pm}}\frac{V'(x)}{g^2(x)}\mathrm{d}x, \qquad (5.3.16)$$

$$r_{\pm}=\int_{x_0}^{x_{\pm}}V'(x)\left(\frac{V'(x)}{g(x)}\right)^{1/2}\mathrm{d}x, \qquad (5.3.17)$$

$$F(w)=\frac{1}{2}(1+w^2\tau^2),\quad F(0)=\frac{1}{2},\quad F''(0)=\tau^2。 \qquad (5.3.18)$$

当取 $g(x)=1$ 时(5.3.15)式就回到文献[4]中的(5.3.5)式，因此我们把加性噪声推广到乘性噪声，且得到一些新的很有意义的结果。从(5.3.15)式不仅可看到有色噪声($\tau\neq0,F''(0)\neq0$)可以导致扩散流，且在白噪声($\tau=0$)时，只要 $\Delta V(x_+)\neq\Delta V(x_-)$时仍然存在着流，这与加性噪声不同，在那里 $\Delta V(x_+)=\Delta V(x_-)$，所以当 $\tau=0$ 时，$S=0$。对于乘性噪声，一般情况 $\Delta V(x_+)\neq\Delta V(x_-)$，这相当于使左、右两边位势峰的等效高度变得不一样，所以即便是白噪声，甚至还是对称周期位势，还都能有定向净流。其次，在加性噪声时流的方向完全决定 $r_{\pm}$ $F''(0)$的值，特别是 $F''(0)$的符号(一般谱可取正或负号)，但在乘性噪声时 $\Delta V(x_{\pm})$值会起重要作用。此外，对于加性噪声，当 $F''(0)>0$ 时，对于给定噪声强度 D 和位势形状，流有一极大值，其相应的 $F''(0)$值为：

$$F_m'(0)=\frac{D\ln(r_-/r_+)}{r_--r_+},\quad F_m'(0)>0, \qquad (5.3.19)$$

即，对于 $F''(0)>0$，$F_m''(0)$是最佳噪声颜色，而在乘性情况，我们将得到一个 $F_m''(0)$的修正公式：

$$F_m''(0)=\frac{D\ln(r_-/r_+)}{r_--r_+}-\frac{-2F(0)}{r_--r_+}(\Delta V(x_-)-\Delta V(x_+)), \qquad (5.3.20)$$

$g(x)$在这里使得 $\Delta V(x_-)\neq\Delta V(x_+)$，因此对给定 D 与位势形状，$g(x)$的形状也是至关重要的，甚至可以使第二项为负值并超过第一项，使 $F_m''(0)<0$，这时流就不存在极大值，即不存在最佳噪声颜色，随 $F''(0)$可能是单调变化的。

为了进一步理解与分析乘性噪声 $g(x)$使得产生一个等效位势，其高度与形状较原来的位势都有所改变，从而流的情况也与加性噪声大

不相同。让我们回忆一下一个布朗粒子在对称周期位势 $f(x)$ 和一个外场 Fx 中的作用，其中 F 是外力（设为常数），我们知道如果外力不存在（$F=0$），布朗粒子向左、右两边对称地扩散，因而飘移速度为零，即无净流，但加上外力 F 之后，周期位势整体随 x 发生倾斜，斜率为 F，因此布朗粒子沿力方向扩散产生定向流，在大阻尼近似情况下，Langevin 方程这时可写为：

$$\nu \dot{x}=F-f'(x)+\Gamma(t), \tag{5.3.21}$$

其对应 Smoluchowski 方程是：

$$\frac{\partial W}{\partial t}=\frac{\partial}{\partial x}\left(f'-F+\Theta\frac{\partial}{\partial x}\right)W=-\frac{\partial S}{\partial x}, \tag{5.3.22}$$

这里 $\Gamma(t)$ 是白噪声，$\langle\Gamma(t)\tau(t')\rangle=2\Theta\delta(t-t')$，$\Theta=kT$，$S$ 为概率流，当温度不太高时，我们可以得到一个定态解，从而计算出飘移速度和概率流，详细推导可以参阅有关文献[12]，我们在这里仅给出最后结果以便与乘性噪声情况进行比较与讨论，概率流 S 的最后表达式是：

$$S=\frac{\Theta(1-e^{-2\pi F/\Theta})}{\int_0^{2\pi}e^{V(x)/\Theta}dx\int_0^{2\pi}e^{-V(x)/\Theta}dx-(1-e^{-2\pi F/\Theta})\int_0^{2\pi}e^{-V(x)/\Theta}\int_0^{x}e^{V(x')/\Theta}dx'dx}, \tag{5.3.23}$$

这里 $V(x)=f(x)-Fx$，周期为 2π，$f(x+2\pi)=f(x)$。从(5.3.23)式可以明显看出，当 $F=0$ 时，$S=0$。

现回到乘性噪声，方程(5.3.3)，先考虑白噪声这一简单情况，即

$$\dot{x}=-V'(x)+g(x)\Gamma(t) \tag{5.3.24}$$

其相应的 Smoluchowski 方程为：

$$\frac{\partial W(x,t)}{\partial t}=\frac{\partial}{\partial x}[V'(x)W(x,t)]+\frac{\partial}{\partial t}\left\{g(x)\frac{\partial}{\partial x}[g(x)\Theta W(x,t)]\right\}, \tag{5.3.25}$$

现假定 $V(x)$ 与 $g(x)$ 为具有相同周期 $L=x_+-x_-$ 的函数，即 $V(x+nL)=V(x)$，$g(x+nL)=g(x)$。我们可以用与上述例子类似的方法，利用周期边界条件 $W(x+nL)=W(x)$ 和归一化条件 $\int_{x_-}^{x_+}W(x)dx=1$，得到一个定态概率流的解析表示式：

$$S=\frac{\textcircled{H}(1-\exp(-\Delta(x_+))}{\left[\int_{x_-}^{x_+}\mathrm{d}x\,\frac{\exp(-\Delta(x))}{g(x)}\int_{x_-}^{x_-}\frac{\exp(\Delta(x))}{g(x)}-(1-e^{-\Delta(x_+)})\right.\\ \left.\cdot\int_{x_-}^{x_+}\mathrm{d}x\,\frac{\exp(\Delta(x))}{g(x)}\int_{x_-}^{x_+}\mathrm{d}x'\frac{\exp(-\Delta(x'))}{g(x')}\right]},\tag{5.3.26}$$

这里

$$\Delta(x)=\int_{x_-}^{x}\mathrm{d}x'\,\frac{V'(x')}{\textcircled{II}g^2(x')}。\tag{5.3.27}$$

可以证明(5.3.26)式分母总是大于零，因此流的方向完全由 $\Delta(x_+)$决定，如果与(5.3.23)比较立即可看到 $\Delta(x_+)$与 F 起着完全相同作用，若 $g=\mathrm{const}$，即加性情况，$\Delta(x_+)\propto\int_{x_-}^{x_+}V'(x')\mathrm{d}x=0$，因 $V(x_+)=V(x_-)$，则 $S=0$，但在乘性噪声情况 $\Delta(x_+)\neq0$，$S\neq0$。

这个问题可以做如下理解：等效力$\frac{V'(x)}{g^2(x)}$可以写为

$$\frac{V'(x)}{g^2(x)}=V'(x)-\left(V'(x)-\frac{V'(x)}{g^2(x)}\right),\tag{5.3.28}$$

其中 $V'(x)-\frac{V'(x)}{g^2(x)}$可以看作附加在周期力的外力，使原来周期力场有所改变，由于在一周期长度上做积分时只留下$\Delta(x_+)=\int_{x_-}^{x_+}\mathrm{d}x\,\frac{V'(x')}{g^2(x)}$，它与零偏离的大小完全决定于 $g(x)$ 的性质，有关 $g(x)$ 与等效位势之间的关系以及推广到有色噪声的推导和结果等将在另外文章中讨论。下面我们着重讨论一个简单模型，即一个过阻尼的布朗粒子在一个时、空周期位势和加性色噪声驱动下的运动情况，用以解释实验中发现的单个动蛋白沿着微管作阶梯式跳跃运动现象。这个模型可认为是由一个不仅与空间，而且与时间都有关的乘性噪声通过适当坐标变换得到的等效方程。

2. *在时、空周期场及加性色噪声中的布朗粒子运动*

在过阻尼近似下，Langevin 方程可写为：

$$\dot{x}=-\frac{\partial}{\partial x}[V(x)-xF(t)]+\varepsilon(t),\tag{5.3.29}$$

其中锯齿型周期场 $V(x)=-k^{-1}\left[\sin(2\pi x)+\frac{1}{4}\sin(4\pi x)\right]$具有光滑边，

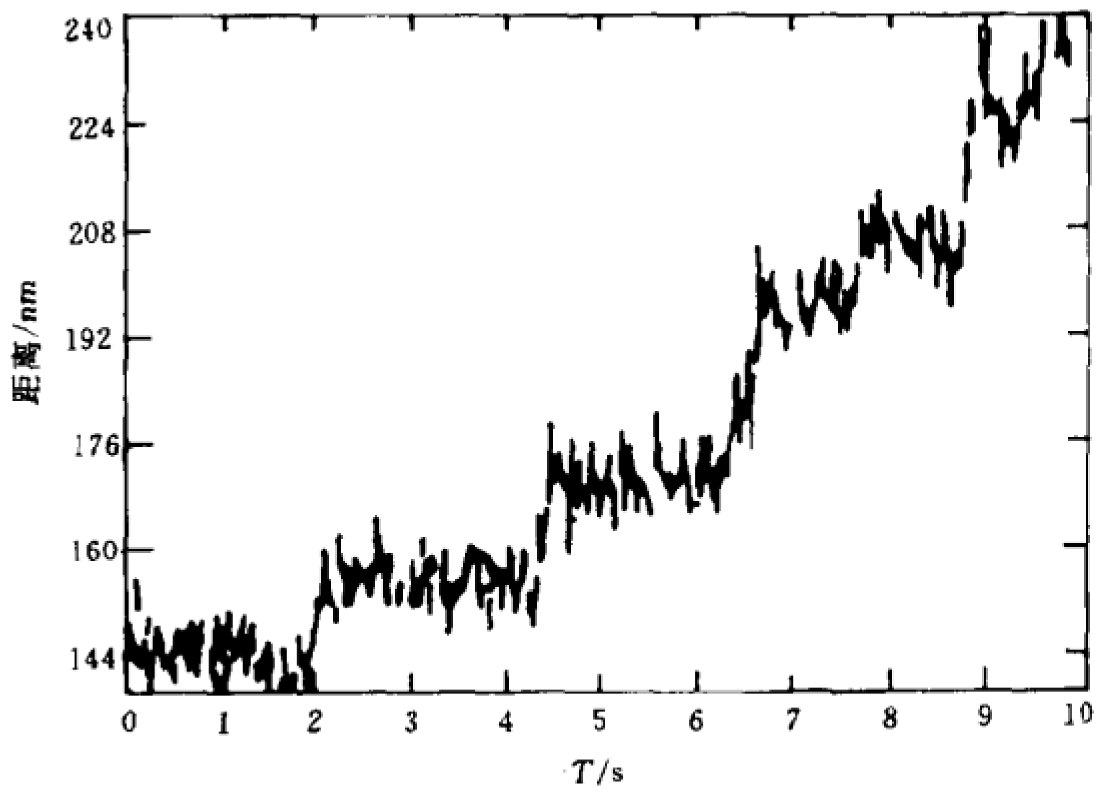


图 5.3.2 动蛋白沿微管作梯跳式单向运动的实验结果(引自文献[2])

周期力 $F(t)=A\sin\omega t$,噪声是关联时间为 τ 的高斯指数型色噪声,满足式(5.3.4)与(5.3.5),$V(x)$的形式见图 5.3.1,$V(x,t)=V(x)-xF(t)$,我们设想可以用这一简单模型来解释单个动蛋白在微管上做周期性跳跃运动现象(图 5.3.2)。首先让我们分析一下物理图象,由于 $V(x,t)$随时间周期地向左、右两边倾斜,而色噪声总是驱使布朗粒子沿位势陡度较小那边方向运动,所以当这一边翘起来时,由位势和噪声所引发的流相互抵消,使得净流很小,甚至为零。反之,当位势陡度小的这一边下斜时,则由位势与色噪声所驱动的流相加,从而导致扩散流猛然增加,这样我们可以期待布朗粒子随时间作跳跃式的扩散运动。为了进一步考查上述图象的合理性,我们用二阶单步叠代

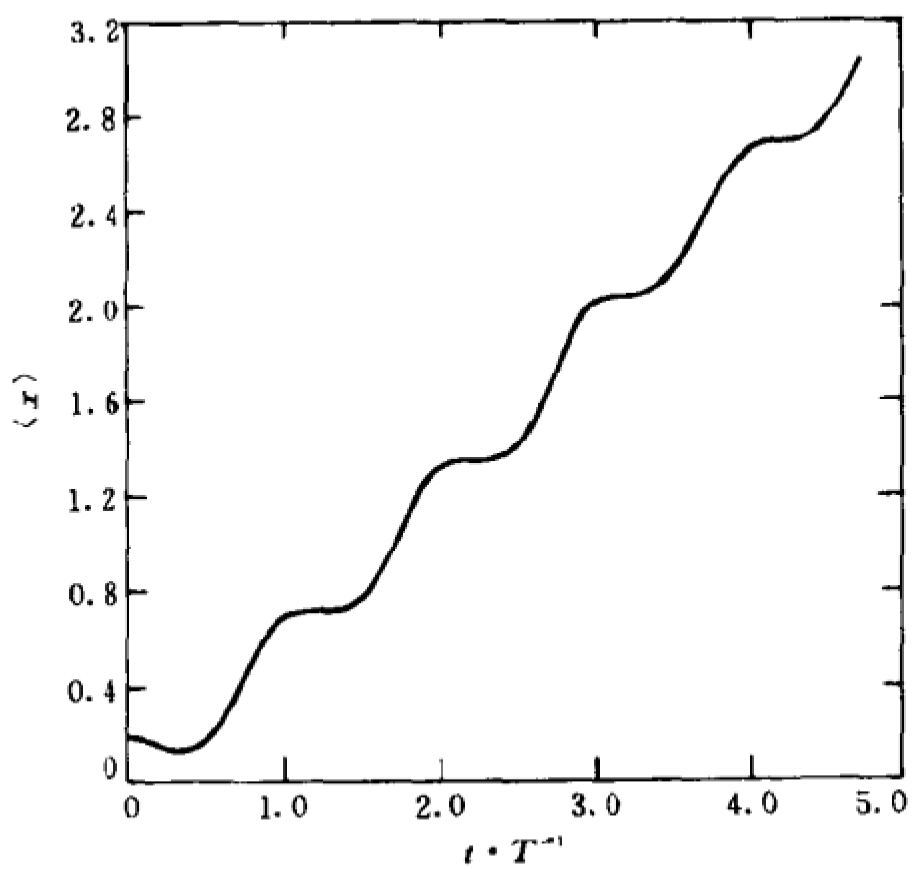


图 5.3.3 对一简单模型进行蒙特卡罗数值计算所得到的布朗粒子位移随时间变化的结果

蒙特卡罗方法[13]，求解 Langevin 方程(5.3.29)，得到位移随时间变化的阶梯跳跃式的运动(图 5.3.3)，与实验结果定性符合，这里所用的参数是 $\tau=0.1$，$D=2.0$，$A=0.5$，$\omega=2\pi$，$T=1$，全部用的是无量纲的量，我们相信通过调参数可以做到与实验在定量上的符合，由于目前模型较简单，主要考虑物理图象上是否合理，对定量上是否完全符合则暂不想作过多考虑。

回顾当前有关分子马达运动的实验与理论状况之后，假定来自 ATP 的噪声不仅是有色的且与其结合位置、水解时间等都有关系，因而采用了乘性噪声数学模型。在定态与弱噪声近似下得到扩散流的解析表达式，并分析了与加性噪声相比的新特点，其中最重要的是乘性噪声等效于加性噪声加上对原周期位势的修正，产生一倾斜的新周期位势。因此即使是在白噪声与对称周期场条件下，也可产生定向运动。为了进一步与实验比较，根据上述等效位势概念，我们考虑一个位势斜率是时间周期函数的简单模型，解释了一个布朗粒子位移随时间呈梯跳式运动现象。

当然这只是很初步的工作，模型还需由实验进一步做检验，一方面要与更多观测量，如力-速度曲线等，进行更定量的比较，另一方面要把模型更加完善化，对模型的基础进行探讨。到目前为止在理论上乘性噪声工作做得不多，若能应用于分子马达这类生物物理问题上去，与实验相结合，反过来将对理论工作有很大促进作用。

参考文献

[1] Huxley A F. *Prog Biophys Biophys Chem*, 1957, **7**: 255; Huxley A F, Simmons R M. *Nature*, 1971, **233**: 533

[2] Svobada K, Schmidt CF, Schnapp B J, Block S M. *Nature*, 1993, **365**: 721

[3] Magnasco M O. *Phys Rev Lett*, 1993, **71**: 1477; 1994, **72**: 2656

[4] Millionas M M, Dykman M I. *Phys Lett A*, 1994, **185**: 65

[5] Astumian R D, Bier M. *Phys Rev Lett*, 1994, **72**: 1766

[6] Prost J, Chauwin J, Peliti L, Ajdari A. *Phys Rev Lett*, 1994, **72**: 2652

[7] Doering C R, Horslhemke W, Riordan J. *Phys Rev Lett*, 1994, **72**: 2984

[8] Bao J D, Zhuo Y Z, Wu X Z. *Phys Lett A*, 1996, **217**: 241

[9] Gilbert S P, Webb M R, Brune M, Johnson K A. *Nature*, 1995, **373**: 671

[10] Dykman M I. *Phys Rev A*, 1990, **42**: 2020

[11] Ghosh A. *Phys Lett A*, 1994, **187**: 54

[12] Risken H. *The Fokker-Planck Equation*, Springer-Verlag, 1984, **11**

[13] Bao J D, Zhuo Y Z, Wu X Z. *J Stat Phys*, 1992, **66**: 1653

(卓益忠 包景东 吴锡真)

附　　录

生物凝聚态

彭桓武

不久前我为《百科前沿》撰写了一个条目“生物凝聚态”，里面注意到这样一个事实，即物理学家常用结构与性质，而生物学家则常用结构与功能。似乎器件的功能是建立在材料的性质基础上的。这样，便可以问什么性质是生物功能的基础？正像金属能导电是由于价带电子结构具有容易受电场作用发生变化使电子形成电流那样，是不是生物大分子特别是蛋白质的结构具有容易发生电荷转移及构象变化这样一种性质，使为实现某生物功能的那些物理的和化学的反应或过程能够顺利进行。如果是这样，这表示对生物大分子低激发态的研究是重要的。我可能没想清楚，没表达清楚，希望就这问题向大家请教。

生物凝聚态(condensed matter in biology)

凝聚态是任何在一个小体积内有多个原子集聚在一起的物质的通称。例如，固体(包括晶体和非晶态)、液体(包括溶液)、液晶、玻璃体、薄膜(包括表面和界面)、高分子、颗粒和小到纳米尺度的团簇等，都属于凝聚态。生物凝聚态则特指在生物体中出现的各种凝聚态，如蛋白质(包括酶)、核酸、脂质、多糖、水和由它们组成的细胞膜、染色体等等。

研究生物凝聚态，与研究一般的凝聚态一样，可分如下方面进行：①其性质的研究，②其结构的研究，③其性质与其结构关联的研究。一般凝聚态其性质包括力学的、热学的、电磁学的和光学的各种物理性质及其化学性质。对于生物凝聚态，还包括各种生物性质。后者，如从更高一个层次来看，也称为生物功能。

一般凝聚态的结构(结构为其内部粒子运动状态的简称)常分为电子结构与离子结构两部分。①电子结构指电子运动的量子态。根据量子力

学，电子的量子态即能级是分立(即非连续)的。不同能级的电子密度分布不同，或使两原子相吸引(成键)，或相排斥(反键)。②离子结构指原子核的运动状态。由于原子内层电子的运动状态与自由原子中的内层电子的运动状态差别甚微，所以常将内层电子与原子核合并作为离子。除氢离子即质子的运动有时(例如在形成氢键时)须用量子力学处理外，一般离子的运动可用经典力学描述。这样，离子结构便可用离子的平均位置在三维空间的分布以及离子偏离其平均位置的各种集体热运动(振动和转动)来表示。对于生物凝聚态，其结构还是分电子结构和离子结构这两部分。不过其离子结构要比一般的凝聚态复杂，或是多种花样组合在一起，或高一个层次，如复合物、缔合物等。

凝聚态的宏观性质与其内部结构有密切关联，这种认识首先来自对简单的分子和晶体的研究；后来对复杂的凝聚态的研究又不断在充实这种认识。1944 年薛定谔在《生命是什么》书中把这种认识推广到生物凝聚态。那时，凝聚态这个从晶体扩充出来的新名词尚未被物理学者采用，所以薛定谔用晶体这个老名词，但他为包含生物的复杂性引入了一个新形容词“非周期性的”。他设想染色体是一个“非周期性的晶体”(用现在术语即是生物凝聚态)，并将遗传的持久性这种生物性质与晶体的分立电子能级间有较大的能量差这种电子结构关联起来，解释了遗传和突变。这些设想为后来发展起来的分子生物学所证实并具体化。

近年来，研究生物凝聚态的结构与其功能的关联，如酶对底物快速催化的高度选择性问题，抗体对抗原的识别和对后者的多变异的适应问题等，虽有很多进展，但仍不失为生物物理学研究的前沿课题。估计在未来发展中的其他前沿课题中，也不免有许多会牵涉到生物凝聚态的结构、功能，以及两者关联的问题。

八 十 自 述

彭桓武

治学与为人之道

治学与为人并没有一个统一的模式，各人有各人的做法，这里我仅以解剖自己的方式给大家提供一点参考，希望对大家有所帮助。先谈三个具有普遍意义的观点作为开场白。

第一，自然界的知识对人类是有用的。这一点虽然在中国古代不太重视，古籍中也少有记载，可在西方几百年前就十分明确基础科学的作用。例如英国皇家学会 300 年前的成立宣言中就强调了科学知识的作用。

第二，历史知识对人类的发展可资借鉴。历史的借鉴作用却是中国人历来所强调的。例如《史记》作者司马迁和《资治通鉴》作者司马光就都说过类似的话。今天我们说的“古为今用”也是同一个意思。历史是不能重复的，但历史知识有参考借鉴的价值。

第三，人与所有生物一样，存在着个体的差异。生物界里没有完全一样的东西，个体与个体之间总有不同之处。每个人所做的事情各不相同，而社会的发展需要人与人之间的通力协作。中国有句古语“天生我才必有用”，英国也有一句意义相近的古谚语：“It takes all sorts of people to make the world.”我们每个人都既有优点又有缺点，因此我们自己应认真分析自身的优、缺点，从而摆正自己在社会中的地位，扬长避短，才能对社会作出最大的贡献。而对领导者的用人之道来说也就应该知人善任，用人之长，避人之短，才能最大限度地提高效率，避免差错。

治 学

我本人在这一生的学习中有这样一些特点。

首先是主动请教。记得上初三的时候，物理老师讲了一个透镜公式

$1/f=(n-1)(1/R_1-1/R_2)$。当时我以为所知道的物理公式都是由实验得出的，看到竟然有这么复杂的公式，甚觉莫名其妙，我便去问老师。我的物理老师是一位北大毕业的好老师。他并不直接回答我的问题，而是借给我一本北大用的英文版大学物理教科书，指着上面的一幅图对我说，解释就在这里。我一看就清楚了，原来这个公式也有实验基础，就是折射定律，其余都是数学推演，主要是用几何学，而且我自己也做出来了。这件事我一直铭记在心，因为我从中得到很深的教益。一是看到了理论的作用，实际上这是我一生中第一次接触物理理论。理论和实验一结合，即使是很复杂的问题也迎刃而解。二是我的英文就此在初三时期就过关了。再有，我看到了一个典型的启发式教育的积极效果。

后来我到英国留学，师从著名学者玻恩(Max Born)。在阅读一本专著时碰到一个散射公式，是用波动力学求解的。波函数写成入射波与散射波的和，而散射的强度需用波函数的二次式计算，但其中交叉项则被略去，然而这些被略去的项在数量级上同保留下来的项不相上下，似乎照理是不该略去的。我便带着这个疑惑去请教玻恩，他给了我一个很精彩的回答。测量散射波是不能在入射波经过的路途中进行的。入射波虽然通常用平面波描写，但实际上总是有一定宽度的，就同光束的宽度受光阑限制一样。在散射方向的测量处，已在入射波的宽度之外，其入射波的幅度应为零，自然也就不出现交叉项了。换句话说略去交叉项恰恰是反映了物理的实在。他的回答告诉我做理论研究一定要使理论能正确地描写实际情形，就是说要妥善处理理论和实验的关系，这是做学问的一大关键所在。

以上两个例子都给了我很大的收获。但我想如果我不主动发问，即使有再好的老师，即使大师级的人物在身旁也是得不到那么大的收获的。因此我深刻体会到，主动请教是做学问的另一大关键。

其次是要处理好学习的深度与广度、简单和繁杂、近期安排与远期目标之间的关系，以便彼此间互相配合，相得益彰。

当年在清华读书的时候实行的是选课制。我在进清华之前已读过汤姆逊(Thomson)的《科学大纲》、达尔文的《物种起源》等许多商务印书馆出版的书，对于科学的体系已有了大概的了解，并且也能确定哪些课程简单一些，哪些课程复杂一点。于是我入学的时候就将四年的选课表统统拿

来,根据繁简、深度和近期远期之间的关系统一选择安排四年所应修读的课程,制定出整整四年所应达到的目标。自然科学课程中我只选物理和化学深入学习,数学只是去旁听并不选课。另外选了一门比较轻松的课程社会学,本来还想选解剖课,因为我想对生物学来说解剖是至关重要的。不过后来没有做成。我觉得只要你深入地学好一门课,那末推而广之,再学相近的科目就也能达到相当深刻的程度。深与广是学习中相辅相成的两个侧面。大家知道,物理学比化学简单,化学又比生物简单,根据由简到繁循序渐进的原则我是先学物理再学化学。最有趣的是学心理学。我找了个高年级的同学一边散步一边聊天,一边谈心理学。他告诉我心理学有所谓三大学派,我便去借了三本书,一派一本,书看完了,心理学也就算学了。

再有就是学习需要兼容并蓄。这是蔡元培先生早年提倡过的。一方面不同的学科有不同的思维方式,我选学不同的学科就能学到各种思维方式,对之我都很尊重。在学社会学的时候我很尊重个体方法(case method),这就是毛泽东所提倡的"解剖麻雀"的方法,而且后来还把这种方法用到学科学中来,对一位量子力学的创始人狄拉克(Dirac)的论文作系统阅读分析。另一方面同一学科有不同的学派,各有短长,我对各种学派都很尊重,这样我就能集各家之所长,并进而形成自己的风格。我曾经跟薛定谔学了四年波动力学,又花了五年跟玻恩学海森伯的矩阵力学,还自己系统读了狄拉克量子力学论文就是一个例子。当然在广泛包容的同时,作分类比较联系贯通也是必要的。

工 作

在工作方面我比较重视如下几点。

1. 认真务实,强调正确反映实际情况,就像画画写生一样,力求真实。

2. 对工作中所涉及到的各种因素都要逐一分析、推敲其重要性,力求抓住关键,可以忽略的就略去不顾。

3. 多途径思考问题,各种途径要加以分析比较,选择最有希望的一条途径先工作,然后再沿第二条途径工作。这是一种搜索方法,在探索或搜索中不断地调整其优先权重。

4. 将大问题分成小问题研究解决。我曾请教过薛定谔如何做研究，他回答：Divide and Command，即分而制之。这样难点就化开了。其实这句话是古罗马的凯撒大帝说的。实践证明这个方法相当有效，不过要做到这一点，要能将大问题分解开来也是需要相当深厚的功力的。

5. 深刻认识矛盾的普遍性与特殊性的关系。这对从事理论物理研究的人来说尤为关键，做好了就能成为好的理论物理学家。理论是有普遍意义的，反映矛盾的普遍性；而实际问题又都是很具体的，表现矛盾的特殊性。从具体的实际情形出发形成研究课题，再用普遍的原则，例如第一性原理去处理。单就实验本身而言，分析实验是从特殊到一般，而设计实验则是从一般到特殊。这种普遍性与特殊性相结合的过程要反复多次才能深入，这也是符合认识逐渐深入的过程的。在研究具体问题的时候注意用普遍性的理论来指导。这样从一个研究对象转到另一个研究对象就不困难，因为总有共同性存在。记得当年我们开始研制氢弹的时候就是如此。当时根本不知道氢弹是个什么样子，谁也没见过，只能从普遍性着手，最终获得成功。这就可以看出认识论中的这一方法是多么有威力。又如我刚开始搞生物物理方面研究时，到上海找人讨教什么是生物物理，但众说纷纭，莫衷一是，连个明确的定义也没有。我便根据普遍性与特殊性的关系分析，生物物理是交叉学科，类似的有化学物理，也是交叉学科，而化学物理的涵义是早就清楚的，杂志都出了几十年了。我从化学物理中看出交叉学科应当有的共性，将它用到生物物理中来。我认为化学物理所以搞得好是因为物理学发展到了一定阶段建立了量子力学，从而从根本上沟通了物理和化学，给化学物理的发展开辟了极光明的道路。因此我认为目前还有待于物理学进一步深入发展到一个新的阶段才能同生物学紧密联系起来，也才能促进生物物理有大的发展。

为 人

以下我想谈谈为人处世当中的一个重要问题，即人生目标的选择与向着目标前进的动力，实际上目标与动力本身也是密切相关的。

我这里讲的目标不是那种所谓宏伟的理想，伟大的抱负，而是人生旅途中在不太长的时间内可以达到的里程碑。因此根据实际情况及时调整目标很重要，这种自觉的调整往往会构成人生的重要转折。

所谓实际情况包括个人的主观条件，即自己的长处和短处，以及客观条件，包括自然的和社会的环境。我自幼体弱，独立生活能力比较差，不能离开家乡到外地上学。于是只能上小学，小学完了也只好再上小学，甚至私塾。早先我上学时，班上一共 30 多个人，我排名 20 多，因为图画、唱歌、手工和体育 4 门课我都不行。虽然算术、语文、英文好，可一平均，成绩就被拉下来了。后来在哥哥帮助下到吉林的毓文中学跳级上初三。在那里没有我不行的课。那里重数、理、化，还能选读解析几何、旧制师范英文等课程。这正好发挥我的特长，一下子就跳到班上第一名。我腾飞了，成了龙头老大。如果仍在原来的学校里老是第二十几名，人被弄得灰溜溜的，也就不会有以后的发展了。因此这是一次对我的一生有深远影响的转折。第二次转折是升大学。我进了清华之后老老实实念了六年书，四年本科，两年研究生，同时身体也变好了。于是我就在"七七事变"之前离开清华，先到泰山避暑，然后转道上海、海防去昆明云南大学教书。看来这是我处理得相当成功的一次转折，以至从不夸奖我的父亲也说我办得好。不久由于西南联大到了昆明，增设了一个出国考场，我也就此考出去了。当时因为剑桥已有好几位中国学生，周培源先生便叫我去爱丁堡师从玻恩。回国后又到昆明教书，我觉得教书总是对社会对国家的贡献。

个人前进的动力有大有小，也有各种各样的动力，难以一概而论。在许多情形下为求生存就是最基本的动力。例如读本科时希望毕业后能做中学老师，读研究生时希望毕业后能到大学任教，这都是指望能有比较好的生活条件，不至为冻馁所困。而在求学的时候，往往动力来自兴趣，来自对自然奥秘的追求探索。在上初中之前我读了大量文、史方面的古书，实际上走的是旧文人史学研究的道路。初三起才接触物理，一下子就被吸引住了，以至"弃文就理"，对自然科学产生了浓厚兴趣，并转化为强大的动力，而且形成了专攻难题的爱好。记得那时数学试卷上有两类题，一类是普通题，每题 20 分，共 5 题；另一类为难题，共两道，每题 60 分。我总是只做难题，也能得 100 分。解出难题带给我巨大的乐趣，又反过来增强我探求自然的动力。当然最强大最根本，最具持久性的动力是与社会责任感密切联系在一起的。新中国成立后组建近代物理研究所，搞原子能，这样自己的工作就同国家的命运联系在一起。后来又发生苏联撤走专家事件，于是国家的安全、荣辱就化成了自我激励的动力，

这就使自己的动力升华到了一个前所未有的高度，远远超过以前的动力。

回顾自己一生我深深体会到，该干的一定得干，而且要尽心尽力地干好。工作越难成长愈多，工作越紧成长愈快。有时会碰到难得的机遇，一定要及时抓住这种天赐良机把工作做得更好。我的经验总结起来可归纳成 16 字诀：

主动继承，放开拓创，实事求是，后来居上。

最后摘录过去写的两首诗中的残联作为结束：

世乱驱人全气节，天殷嘱我重斯文。(1939，爱丁堡)

不是工农兵协力，焉能数理化成功。(1964，罗布泊)

个人学术思想漫谈

今天我在这里同大家漫谈，以随想的形式回顾多年来在学术领域中我所接触到的几个问题，以及我自己做过的一些工作，讲一讲我的看法。

关于广义相对论

早在读大学本科的时候，我就开始了广义相对论方面的研究。我在四年级时的毕业论文就是做广义相对论，由周培源先生指导。后来周先生又招我做研究生，继续广义相对论方面的研究。不过周先生将题目交代给我之后就去了美国爱因斯坦那里，留下我独自在清华。由于研究生院头两年要听许多课，我只做了一些数学方面的工作，不过直到我离开清华时都未发表，事实上当时国内局势也无法发表，大家都逃难去了。多年以后从国外回来才知道北欧有个克莱因(F. Klein)教授发表了类似的工作。当年周先生是按爱因斯坦的观点教我们的，当时我总觉得不舒服，以至于不知道广义相对论是个什么理论，是参考系理论呢还是时间、空间理论，抑或引力理论。其实周先生虽用爱因斯坦的观点教我们，但已对爱因斯坦的观点产生了疑问，酝酿提出新的解释。

直到 70 年代，周先生在庐山物理学年会期间正式提出与爱因斯坦不同的观点，将谐和条件作为物理条件。这是一个很大的变化，因为这样一来就没有广义相对论的弯曲时空，只有狭义相对论的闵可夫斯基时空或伽里略时空。参考系理论仍是通常的洛伦兹变换理论。周先生的观点告诉我们做研究要独立思考，实事求是，即使像广义相对论这样的著名学

说，只要发现有问题仍然可以讨论。当然也不只是周先生一人对广义相对论提出质疑。前苏联的福克也持类似的观点，只是不如周先生那么明确。他们认为谐和条件是可以人为加上去的，而周先生认为必须用谐和条件，这是和爱因斯坦引力方程同样重要的，是引力理论的一部分。福克也认为广义相对论中没有参考系理论。从一个参考系到另一个参考系的变换可以用群论方法来数。6 个转动，4 个平动，组成 10 个参数的群。根据微分几何中的定理，n 维空间可能允许的最大的群为 $n(n+1)/2$，$n=4$ 时这个数目为 10。因此必定有空间的曲率为常数。曲率若为零就是狭义相对论，若曲率不为零则可以有另一个空间。因此，作为参考系的理论，广义相对论的提法是不通的。在周先生的理论中时间空间就是狭义相对论的 g_{uv}，是引力势，引力场有 10 个势。从而说明广义相对论只是引力理论。当年我第一次到英国时曾问过玻恩(Max Born)，爱因斯坦为何要搞广义相对论，回答是狭义相对论建立起来以后就一定要推广牛顿的引力理论。标量势不行，光不能拐弯，与实验不符，而矢量势要排斥，不能相吸。如要相吸就必须搞到有 10 个分量的二级张量去。我注意到根据周老观点狭义相对论的时空观是同量子力学、量子场论的时空观一致的，这就避免了彼此无共同语言的局面。另一方面周老的理论也不是完全否定爱因斯坦，另起炉灶，仍是爱因斯坦的引力场理论。只是 10 个引力势所满足的 10 个方程彼此并非完全独立，其中有 4 个恒等式，必须加上 4 个谐和条件作为物理条件方可定解。我觉得根据这个观点可以开展许多联系实际的问题的研究。研究爱因斯坦广义相对论的都是数学家，追求所谓严格解。可这些严格解一个也不同物理问题相对应，因为不知如何同物理问题相对应。然而根据周老的观点就可以做与物理相对应的问题。譬如对一个密度为常数的球的问题。按照爱因斯坦的引力方程，坐标可以任意选择，只要应用史瓦西(Schwarzschile)坐标即可求解。但史瓦西坐标不知是什么意思。而按周老的观点，既然是狭义时空，坐标是客观的。10 个引力方程加上 4 个谐和条件，解 14 个方程才能得到唯一解。从这可以看出采用周老的观点接触实际问题是可能做出结果来的。这样就打开了理论联系实际之门。

关于晶格动力学和量子化学

我到玻恩处的时候，已有一位师兄在那里，他在用量子力学做凝聚态

研究方面很有名气。玻恩和他商量给我出个题目，就是计算金属原子热振动的频率，而且建议我用弗勒利赫(Froehlich)边界条件和微扰法做。我一看就发现一级微扰好做，既简单又漂亮。由于简谐运动是位移的二次函数，位移就是微扰变量，因而计算频率一定要做二级微扰，但做二级微扰相当困难。我采用另外的方法，即作一变换。在变换空间里恢复了晶格的周期性。当然动能、位能和算符都变了，但由于恢复了周期性的边界条件，因而采用正规的微扰理论做到二阶并不困难。不过由于当时的哈特里-福克(Hartree - Fock)近似不能得出关联能，因而振动频率做不准，只求得了弹性模量。所以我的博士论文可算只做了一半。但我以后一直关注这个问题。前些时我终于找到了可以计算关联能的框架。从 1940 年交博士论文直到现在 90 年代，这个问题在头脑里呆了半个多世纪。可见有的研究课题够你想一辈子的。关于周期性边界条件值得一提的是拉曼(Raman)和玻恩就此有过很激烈的争论。当时似乎是坚持这一边界条件的玻恩占了上风，因为理论的结果很好。拉曼从分子的观点出发认为任何实际的晶体都不存在周期性边界条件。我觉得两种观点都有可取之处，不应厚此薄彼，就像搞金属，量子力学可以采用布洛赫(Bloch)函数也可以采用旺尼尔(Wannier)函数一样。

后来我到爱尔兰的都柏林听海特勒(Heitler)讲量子化学课。我对量子化学一直有兴趣，因为我也学过不少化学。但我对量子化学也是一直不满意。因为量子化学是物理学家做的，而物理学家并不知道化学的真正要求所在。就拿化学反应中键能的变化来说，量子化学不管分子的大小，一概当成一个物理问题来计算电子的能量，直到原子内部 K 壳层的能量都一视同仁。于是一个电子的能量就高达上百电子伏，可化学键能不过在电子伏的数量级，在化学反应中键能变换的差别就更小了，还不到 1 电子伏，结果是每个电子能量都算出来相加，再总能量相减，变成大数相减，其误差便可想而知。再有，基函数选得如何连自己也不知道精确度怎样估计，又引进了误差因素。因此这样算出来的很多东西都是不可靠的。实际上这方面只要计算键能部分，计算差别，而内层电子的能量并不重要。因此我曾建议用键函数，而不用每个电子的波函数。当时想做的是 CH_4，化学家称之为标准状态。只要算出 C—H 与 C—Cl 两个键即可算出五个分子，即 CH_nCl_{4-n}($n=0,1,2,3,4$)，还可以自洽核对，自然误差就小。求得

键波函数，键-键相互作用还可扩展开去深入研究。这是我转到搞原子能之前最后一位研究生的工作，可惜并没有完成。前几年听说有人搞量子化学要采用双电子波函数，其实键函数就已是双电子波函数。

量子化学的缺点说明不能用单纯的物理观点搞化学问题，不同的问题应当有不同的观点。目前的量子化学其实只是物理。我认为应当将量子力学同物理化学结合起来，建立量子物理化学。例如电负性的问题，定义有好几个，这个问题搞清楚了就是量子物理化学。

我自己搞得比较深入的是场，电磁场、介子场。场是无穷个粒子的量子力学，有无穷个自由度。给定场，作傅里叶变换，系数即坐标。无穷个粒子就有无穷个坐标，因而引起发散的困难，我也为之花过很多精力。海特勒提出不必理会发散问题，保留项即为阻尼效应。我曾试用阻尼理论解决介子场的发散困难，不过没解决电磁场的发散困难，因为收敛不了，还是欠了帐。我现在想用做关联能的方法再做一下。老一辈的物理学家，如狄拉克，大都不满意重整化理论，因为数学不干净，不严格。我也不满意。我现在想追求的是数学严格的方法。

关于应用数学

应用数学必须解决实际问题。我回国不久就开始发展原子能，需要做许多数学工作。反应堆，中子扩散中的方程类似热传导方程，更严格一些则为玻尔兹曼方程。但我做的第一个可与实验比较的问题却是在炼钢方面，是关于高温钢锭快速加热的问题。这是鞍钢为推广苏联的经验而提出来的，目标是钢锭从炉中取出后不让炉温下降就放进新钢锭，从而节省时间、提高生产率。因此必须研究适用于哪些钢种，钢锭的尺寸要多大等。当时组织了一个班子，除鞍钢、抚钢和北京钢铁研究院的人，还有我和北大的王竹溪教授以及钢铁学院的教授，一起进行实验。主要观察加热钢锭剖开后有无开裂、破坏。同时我们应用热传导、弹性力学的知识从理论上计算，可以考察某一参数(例如弹性模量，钢锭尺寸)在各种数值下的影响，而且不花钱。结果发现这个苏联经验适用于碳钢。而特殊合金钢就不能采用。最后由有色金属局一纸命令下了结论，从而完成了任务。对我来说更重要的是我的计算同实验结果成系统性的相符，增强了我的自信心。接下来搞的一项实际工作才同原子能有关。在核工厂里造浓缩铀，要把许多

叫做扩散基的设备连接起来。这是化工方面的事，我也懂一些。那种工厂里处理的都是可裂变的危险材料。如果处置不妥就会出链式反应事故。我去的任务就是要保证工厂不出事故。核工业的安全要做到亿万分之一的程度，要求是很高的。我到那里之后对各种设备逐一仔细查问工厂里的工程师，看哪个环节可能会出毛病。主要原料六氟化铀是气体，运输时要腐蚀机械，造成冷却水的泄漏，冷却水又会同气体起反应。诸如此类的问题举不胜举。我花了整整三天搞调查。一层一层地反复考虑各种各样的可能性。做这一类实际问题要求很广的知识面。普通物理、普通化学和普通工程学方面的知识都要用到。诸如应力、腐蚀、蚀孔口径与流量的关系、气体化学反应的速度等等。所有这些都要进行估算并提出监测办法，选择监控点的位置，哪里最容易出危险就应把监控点设在那里。最终形成一套规章制度。三天调查后我开了一个通宵的夜车拿出了方案。当然在这整个过程中都要用到应用数学。

关于交叉学科

我从核武器战线下来后觉得自己很合适搞交叉学科的研究，因为理与工两个方面我都有比较广的知识面，虽然哪一门都没有能很深入。交叉学科兼具普遍性与特殊性。我开始时选择了生物物理。为此我特地跑到上海生物研究所，请教什么是生物物理，但没有人能说出定义，在北京也一样。于是只好退一步转向搞化学物理。这门学科比较明确，在美国化学物理杂志都出了几十年了。我搞化学物理的目的是想弄清楚交叉学科的真正涵义以及如何进行交叉学科的研究。另一方面，研究化学物理也对生物物理有用，因为生物与物理都同化学密切相关。后来我理解到生物物理起码应包括生物凝聚态（这是我给起的名字，以前称生物聚集体）的凝聚态物理以及生物化学的化学物理，也许还包括系统控制和信息论。生物物理有许多特殊的规律。例如统计力学在生物凝聚态中就不一样，不是各态历经的。于是只能取能量面上的一部分求平均。因此统计依赖于取平均的区域，严格讲统计取决于在何处取平均。结果便是有个体差异，分类差异。生物就是有差异的。例如调查人群的血型可以预测下一代不同血型间的比例，不同区域的情况就不同。

化学反应最大的特点在于从普通化学里学到的那些反应是根本不存

在的。实际上反应都有好几步,8 步,10 步,称作基本化学步骤。基本的意思是只牵涉少数几根化学键,两三个原子。普通化学是将许多反应加在一起算总帐。许多基本反应平行地或串联式地发生,都是态-态反应,涉及到量子态。有时反应结果得出布居数反转,即可产生化学激光。这里不是热化学反应。反应物与生成物都不处于热平衡。问题牵涉到量子截面的大小,但截面分布并不符合玻尔兹曼因子,称之为出奇,也可称为负熵。因此,热化学中熵一类的问题就值得研究。因为都是非平衡态,如何有熵的定义。不过如果在某个步骤予以截断,使之处于平衡态,还是可以有熵的。总而言之有许多有趣的问题需要严格的脑袋去思考。这一类反应往往表现出目的性,例如某个产物出来了就会抑制那个反应,便变成为毒物,这就表现出事后的目的性。

很显然,个体的差异性和目的性是生物物理的两大特点。但我的理解还是很粗浅。

回顾 60 多年来从事科学研究的经历,对象千变万化,工作一时告个段落,学术思想的演化则需要较长时间。纯粹科学与应用科学的共同发展,融真善美于一体,更需要群众集体的努力。我深感科学事业值得我们穷毕生的精力愉快地去追求、去探索,"精诚求实毕生愉,与善美真为侣"。马克思的名言:"只有那些不畏劳苦,沿着陡峭山路攀登的人,才有希望达到光辉的顶点",应当永远成为鼓舞我们前进的强大动力。

彭桓武先生科研教育工作60年

黄祖洽

为了祝贺彭桓武先生八十华诞，我写了一篇文章(以下简称“祝贺”)，在《物理》杂志1995年10月号上发表。另外，这年3月彭先生应邀访问复旦大学期间，在物理系作了两次长篇演讲，回顾了自己的治学、为人之道和个人的学术思想，已经由该系蒋平同志整理成文，在《科学》杂志上发表(1996年第2期)。该文也收入了本文集，见彭桓武：“八十自述”，以下简称“自述”。所以我今天只根据彭先生的几首诗词，着重介绍我所了解的彭先生60年来科研和教育工作的一些情况。

记得10年前，在度过古稀之年的时候，彭先生写过一首“七十自况”的七律：

少小离家自学飞，省垣故国识芳薇，
华园六载登堂座，云海多年入室帏。
众木喜看撑大厦，群禽协舞映朝辉，
一场淫雨风滥后，韧翼总凭余热挥。

在他这首诗的自注中写道：“前四句记叙解放前求学经历(省垣指今吉林市，故国指今北京市，华园指清华大学及研究院，云指云南昆明，海指海外爱丁堡及都柏林)；后四句则概括解放后参加祖国建设的形象。”

实际上，从1935年毕业于清华大学物理系后作研究生，跟周培源教授进行相对论的研究开始到现在，彭桓武先生从事理论物理研究和教学工作已经有了整整60年的时间。这60年中，他先后在相对论、固体理论、介子物理、量子场论、核物理以及与原子能科学技术有关的物理理论等研究领域中取得了令人注目的成就；与此同时，彭先生也总是在不同场合，通过讲课、带研究生或年轻人作研究等不同形式，进行着教育和培养人才的工作。

1937年7月7日，抗日战争爆发，清华大学南迁。彭桓武辗转去昆明，在云南大学理化系任教员。1938年秋，他考留英庚款被录取后赴英国，入爱丁堡大学理论物理系，随著名物理学家玻恩(M. Born)作研究

生。1940 年以固体理论方面的论文“电子的量子理论对金属的力学及热学性质之应用”，获哲学博士学位。这方面他和 K. Fuchs 合作的代表性论文“Crystal theory of metals: Calculation of the elastic constants”发表在 *Proc. Roy. Soc.*，London，A180 (1942)上。1941～1943 年彭桓武继续在爱尔兰的都柏林高等研究所作博士后的研究学者(Scholarship)，和海特勒(W. Heitler)合作进行介子理论方面的研究，发展了量子跃迁理论，用以处理核碰撞中产生介子的过程，得出了能谱强度，并根据它首次解释了宇宙线的能量分布和空间分布。这就是当时名扬国际物理学界的，以作者海密尔顿、海特勒、彭(Hamilton，Heitler，Peng)三人姓氏缩写为代号的，关于介子的 HHP 理论。在此理论中已经出现了后来被称为 Dyson 方程的方程。他们的代表性论文“Theory of cosmic - ray mesons”发表在 *Phys. Rev.* 64(1943)上。从该文的摘要中可以看到，文章是应用后两作者所发展的量子阻尼理论作出的。1943 年 7 月，他在英国爱丁堡大学作博士后的卡内基研究员(Carnegie fellowship)，和玻恩等合作进行场论方面的研究。1945 年以场论方面的论文“量子场论的发散困难及辐射反作用的严格论述”获得爱丁堡大学科学博士学位后，又到都柏林高等研究所，在由波动力学创始人薛定谔(E. Schrodinger)任所长的理论物理研究所任助理教授两年，继续作场论中用生成函数方法表示波函数的研究工作，并为了和实验室中介子的人工产生作比较，具体指导来自法国的访问学者摩勒特(Cecile Morette)对较低能区核碰撞中的介子产生作了更细致的计算。他先后两度在爱丁堡与都柏林两地从事理论物理研究共计 9 年，至 1947 年底才回国。这 9 年间，彭桓武经常和蜚声国际物理学界的大师玻恩、薛定谔及海特勒等人共同从事上述各方面的研究。在共同研究中，通过自己的努力和同事间的相互启发，他的理论修养、工作能力和对学术问题的见识与判断能力都得到了很大的提高，并且以自己的创造性劳动对共同的工作作出了重要的贡献。后来，海特勒在写到都柏林高等研究所时曾经这样提到彭桓武：

“……同事中最受热爱的一个是中国人彭桓武……，经常的兴致结合着非凡的天才，使他成为同事中最有价值的一个。”

这九年的学习和研究工作对彭桓武先生学术思想的影响，详见前述的“祝贺”和“八十自述”两文，这里就不一一介绍了。

1947 年底，在中华人民共和国成立前夕，彭桓武回到祖国。先在昆明云南大学物理系任教授，开设物性论、高等电磁学和量子力学等三门课，同时开始了关于核力的研究工作。1949 年初，天津、北平相继解放，他便绕道香港，经天津到北平清华大学任教授，先后开设普通物理、量子力学及数学物理方法等课程，并招收理论物理方面的研究生。我有幸从此开始向他学习作学问、作事和作人之道。1952 年 10 月至 1955 年 6 月彭先生在北京大学物理系讲授量子力学，1954 年暑假在青岛由教育部举办的讲习班中为各大学培训了一批量子力学的师资。他讲授量子力学，既注意使学生清晰掌握基本的物理概念，又注意引导学生将量子力学用于解决原子、分子等微观系统中的物理问题，从而培养他们解决问题的能力。他平易近人，和学生亲如朋友，常到学生宿舍谈心，或与学生一起漫步园林，在散步中讨论问题，讲述他的见解。春风化雨，诲人于无形。他的这些见解，我们现在还可以从"八十自述"一文中读到。

1955 年 10 月以前，他的研究工作主要是将量子力学应用于原子核这一多体系统，特别是包含 2～4 个核子的轻核系统，利用有关轻核(^2H，^3H，^3He 及^4He)的基态结合能、^2H 核的虚态能级以及核子-核子散射的周相等有关实验数据，来探索核力的形式和处理核多体问题的方法。在他指导下先后参加这一工作的年轻人有云南大学的唐懋荧(1949 年)，近代物理研究所的金星南和张继恒(1950 年)、黄祖洽(1951 年)，以及他在北京大学的研究生周光召和严肃(1952～1955 年)。这一工作在当时是走在国际前列的。它一方面可以说是彭桓武在国外关于介子问题研究的继续，另一方面也标志着他把注意力转向原子核物理和核能应用的开始。从 1953 年到 1955 年，他还在当时物理所的理论物理室主持了一个核理论的讨论班，集体学习和讨论核物理中的理论问题。在进行核理论和量子化学理论等基础研究的同时，他也注意研究解决工业生产中提出的实际问题。早在 50 年代初期，针对我国最大的钢铁基地——鞍钢为提高生产效率，拟采用快速加热钢锭的新工艺，但必须防止裂缝产生的需要，他就应邀探讨过高温加热中钢锭的安全直径问题。他巧妙估算出的安全直径理论值和实验结果一致，为我国高温热处理第一个规程的制定提供了依据。

他在国外和国内发表了论文 30 余篇，其中在国外发表的有固体理论论文 3 篇，介子理论与场论论文 15 篇；国内发表的有核理论论文 4 篇，理

论物理和数学物理论文10余篇，其中"包含界面条件在内的变分近似法"一文在解决一些实际问题中遇到的活动边界问题中得到了应用。1985年曾由中国学术出版社出版了他的论文选集。该选集目录中也包括了未入选论文的题目。

彭桓武1945年曾因关于场的量子力学和统计力学的一系列探索性工作而与玻恩共同获得英国爱丁堡皇家学会的"Macdougall - Brisbane奖"。1948年被选为皇家爱尔兰科学院院士。1955年6月被选为中国科学院学部委员。1982年，由于在原子弹氢弹设计原理中的物理力学数学理论问题所取得突出成果中作出了重要贡献，他领衔和其他同事一起获得了国家自然科学一等奖。

1955年10月～1956年4月，他参与由钱三强带领的实习团去苏联学习反应堆理论。从此，一直到1972年11月，他的主要精力放在发展我国原子能事业所需要的培养青年干部、理论研究和学术组织工作上。1956年上半年，他和黄祖洽合作，在物理所举办了为期一年的反应堆理论训练班，学员近20人，为中国培养了第一代反应堆理论研究人员。1956年10月～1957年5月在北京大学技术物理系开设反应堆理论与核工原理两门课程，为中国的原子能事业培训了大批青年力量。1962年9月～1964年6月，他在中国科技大学01系讲授流体力学并指导毕业班的补充讨论，使听课的学生在学习原理之外，进一步接触到实际问题的解决途径。

50年代后期和60年代前期，他协助所长钱三强，参考前苏联专家提出过的一些意见和前苏联原子能研究所的学术组织体制，结合当时中国原子能研究所的实际情况，具体指导各研究室明确研究方向和制订计划，并指导所的职能部门进行研究，对全所的科研工作进行了整顿。在反应堆及核动力研究方面，在他的学术指导下，更具体而合理地组织了关于反应堆的理论、实验和工程设计工作的科技队伍，加强了他们彼此之间的协作。为了使我国第一个重水堆在建成启动后不断提高反应堆运行的质量，使它在科研和干部培养上更好地发挥作用，他在1959年组织了一系列的学术讨论会，由与反应堆有关各研究室的人员参加，要求负责反应堆各个系统运行的科技人员提出报告，说明运转中的经验、遇到的问题和改进的意见，并由大家进行讨论。经过这些生动的学术活动，不仅提高了反应堆工作者的水平，引发了他们深入钻研有关业务的兴趣，而且也为以后这个

重水反应堆进行改建提出了方向，奠定了基础。同一期间，他和金星南合作，在该所培养并组建了一支计算数学队伍。作为核工业部核临界安全小组的第一任组长，他亲自参与并带领全组解决了核燃料生产和加工过程中所遇到的一系列临界安全问题。在核武器研制方面，他运用强有力的理论手段把复杂的方程组予以简化，完成了原子弹反应过程的粗估计算，划分了反应过程的各个阶段，提出了决定各阶段反应过程特性的主要物理量，对掌握原子弹反应的基本规律与物理图象起了重要的作用。在探索氢弹理论设计原理的过程中，他更发挥了自己深刻的理论洞察力和民主讨论的学术领导特长，引导更年轻的同志进行物理机制和力学规律等各方面的研究。在解决以上这些重要问题中，彭桓武以自己的科学实践为榜样，引导和培养造就了一批优秀的年轻科学工作者，其中有不少已被选为中国科学院的院士。

1978 年，彭桓武调任中国科学院理论物理研究所所长后，应中国科技大学研究生院的邀请，于同年 10 月至 1979 年 6 月开设理论物理课程。当时来听课的也有北京其他单位的年轻人。1980 年，他鉴于凝聚态物理对我国科学发展和国民经济的重要性，在各种学术讨论会上大力倡导，并参与组织了科学院数理学部的凝聚态理论和统计物理学术小组，被选为该小组的第一任组长，致力于推动这门学科在中国的普及和发展。1982 年 2 月，又在北京大学物理系讲授分子反应动力学，旨在国内提倡化学物理这门新兴的交叉学科。近年来则在有关学术会议（包括这次香山科学讨论会）上多次大力提倡生物物理的研究，就如何开展这方面的工作提出自己的看法。年近八旬的彭桓武先生，在进行这些倡导工作和主持理论物理专项基金的同时，仍然孜孜不倦地亲自动手作一些感兴趣的理论研究，例如多电子强关联系统的处理方法，求解相对论方程时谐和条件的作用等方面的研究。遇到适当的机会，他也和年轻人谈些治学的体会。在一次教师节的谈话中，他把学习方面的经验归结为四句话，即“学问主动（这点最重要），学友互助，良师鼓励，环境健康”。对于教师的作用，他认为主要是对学生的鼓励，最多是鼓励加指导，但最重要的是鼓励，提出题目来，或学生自己找到题目，就鼓励他们自己去钻研，去创造，而不是要求他一定按教师的办法去作，束缚他们。从学生来说，当然要不受别人束缚，更重要的是自己要有实事求是的精神，对问题作独立思考，只有这样，才能有创造，

有发展，后来居上。在“回顾”一文中，彭先生更系统化地说明了这些观点。

1995年5月15日，彭先生写下了一阕“吟两朵白芍药”的西江月词：

紫草黄花点缀，白盘绿叶帮扶，婷婷顾盼若召予，脉脉含情无语。

玉洁冰清帝女，心专目定凡夫，精诚求实毕生愉，与善美真为侣。

他在词后自注说：“白盘指白芍药的脸盘，帝女、娥皇女英指两朵白盘，亦即善美真（基础科学与应用科学）的化身，凡夫指作者即吟花人。”这阕词生动地表达了彭先生60年来对基础科学与应用科学“精诚求实毕生愉”的情怀。

彭先生晚年，生活方面仍坚持自理，乐观旷达，注意锻炼，常常自己或偕同友人去颐和园、北京植物园或香山等处游览，去时总是搭乘公共汽车。常风趣地说，散步可以练腿，乘公共汽车时用手攀伏站着可以练臂，在家有工夫时作作研究可以练脑，而经常作家务劳动则有利于长寿。今年元旦，彭先生自撰了一阕新词——“香山恋”，舒发自己的心情。现在抄下词的后半阕如下：

……

喜迎晚景清贞，诗思萦盈。

廉颇刀利，太公钩直，又几番壮我征程。

愿宁静而致远，求深新以升腾。

惟童心不泯，耄耋期颐总似年轻。

“桃李无言，下自成蹊。”彭先生爱国主义的精神，多作贡献的实践，科学求真的态度，学术民主的方法，奖掖后进的用心，和淡泊名利的胸怀，乐观旷达的性格，都是值得我们钦佩和努力去学习的。值此彭先生80华诞之际，敬祝他健康长寿，童心不泯，壮心不已，“韧翼总凭余热挥”，为国家现代化作出更多贡献。

彭桓武先生著作目录

研究论文

[1] Peng H W. Perturbation Theory for the Self－Cons istent Field. *Proc Roy Soc*, London, 1941, **A178**: 499

[2] Fuchs K, Peng H W. Crystal Theory of Metals: Calculation of the Elastic Constants, *Proc Roy Soc*, London, 1942, **A180**: 451

[3] Heitler W, Peng H W. The Influence of Radiation Damping on the Scattering of Mesons, Part II. *Proc Camb Phil Soc*, 1942, **38**: 296

[4] Peng H W, Power S C. Stability of Rhombohedral Bravais Lattices. *Proc Camb Phil Soc*, 1942, **38**: 67

[5] Heitler W, Peng H W. Anomalous Scattering of Mestons. *Phys Rev*, 1942, **62**: 81

[6] Heitler W, Peng H W. On the Production of Mesons by Proton－Proton Collisions. *Proc Roy Irish Acad*, 1943, **49A**: 101

[7] Hamilton J, Heitler W, Peng H W. Theory of Cosmic－Ray Mesons. *Phys Rev*, 1943, **64**: 78

[8] Hamilton J, Peng H W. On the Production of Mesons by Light Quanta and Related Processes. *Proc Roy Irish Acad*, 1944, **49A**: 197

[9] Peng H W. On the Cascade Production of Mesons. *Proc Roy Irish Acad*, 1944, **49A**: 245

[10] Born M, Peng H W. Quantum Mechanics of Fields, Part I. *Proc Roy Soc*, Edinburgh 1944, **A62**: 40

[11] Born M, Peng H W. Quantum Mechanics of Fields, Part II. *Proc Roy Soc*, Edinburgh 1944, **A62**: 92

[12] Born M, Peng H W. Statistical Mechanics of Fields and the Apeiron. *Nature*, 1944, **153**: 164

[13] Born M, Peng H W. Quantum Mechanics of Fields, Part III. *Proc Roy Soc*, Edinburgh 1944, **A62**: 127

[14] Peng H W. The Divergence Difficulty of Quantized Field Theories. *Nature*, 1944, **154**: 544

[15] Peng H W. On the Divergence Difficulty of Quantized Field Theories and the Rigorous Treatment of Radiation Damping. *Proc Roy Soc*, London, 1946, **A186**: 119

[16] Peng H W. On the Representation of the Wave Function of a Quantized Field by Means of a Generating Function. *Proc Roy Irish Acad*, 1947, **51A**: 113

[17] Morette C, Peng H W. Cross Sections for Production of Artificial Mesons. *Nature*, 1947, **160**: 59

[18] Morette C, Peng H W. On the Production of Mesons by Nucleon－Nucleon Collisions. *Proc Roy Irish Acad*, 1948, **51A**: 217

[19] 彭桓武，唐懋荧. 关于^{2}H，^{3}H，^{3}He，^{4}He诸原子核结合能之问题，第一部. 中国物理学报，1950，**7**: 309; Peng H W, Tang M Y. On the Binding Energies of the Atomic Nuclei ^{2}H,

^{3}H, ^{3}He and ^{4}He, Part I. *Chinese Journal of Physics*, 1950, **7**: 316
[20] 金星南，张继恒，彭桓武. 关于^{2}H，^{3}H，^{3}He，^{4}He诸原子核结合能之问题，第二部. 中国物理学报，1950，**7**：324；King sing－Nan, Chang Chi-Heng, Peng H W. On the Binding Energies of the Atomic Nuclei ^{2}H, ^{3}H, ^{3}He and ^{4}He, Part Ⅱ. *Chinese Journal of Physics*, 1950, **7**：331
[21] 张继恒，彭桓武. 关于^{2}H，^{3}H，^{3}He，^{4}He诸原子核结合能之问题，第三部. 中国物理学报，1950，**7**：339；Chang Chi－Heng, Peng H W. On the Binding Energies of the Atomic Nuclei ^{2}H, ^{3}H, ^{3}He and ^{4}He, Part Ⅲ. *Chinese Journal of Physics*, 1950, **7**：348
[22] 彭桓武，黄祖洽. 应用忽鲁登变分法决定核子与核子散射的周相. 中国物理学报，1951，**8**：150；Peng H W, Huang Tzu－Chia. Determination of the Phase for Nucleon－Nucleon Scattering by Hulthen's Variational Method. *Chinese Journal of Physics*, 1951, **8**：159
[23] 彭桓武. 关于分子结构的一个新的量子力学计算方法. 物理学报，1955，**11**：143
[24] 彭桓武. 寿命关联实验的重要性. 物理学报，1962，**18**：165
[25] 彭桓武. 反应堆理论的变分法形式. 原子能科学技术文献，堆物003，中国科委，1962
[26] 彭桓武. 带圆孔的方柱中的热传导. 物理学报，1963，**19**：268
[27] 彭桓武. 边界形状的变化对偏微分方程本征值的影响. 物理学报，1963，**19**：538
[28] 彭桓武. 处理偏心圆柱形气缝导热的一个微扰方法. 物理学报，1964，**20**：137
[29] 彭桓武. 包含界面条件在内的变分近似法. 科学通报，1975，**20**：416
[30] 彭桓武，庄杰佳. 超高能粒子加速的新途径——调相聚焦激光加速器. 中国科学，1979，**10**：953；Peng H W, Zhuang Jie－Jia. A New Way of Superhigh Energy with a Phase-Adjusted Focusing Laser Accelerator. *Scientia Sinica*, 1980, **23**：159
[31] 彭桓武. 阻尼谐振子的量子力学处理. 物理学报，1980，**29**：1084
[32] 彭桓武. 五维空间中电子的流体结构尝试. 物理学报，1981，**30**：1030
[33] 彭桓武. 高温加热中钢锭的安全直径. 见：王竹溪等编. 理论物理论文集. 科学出版社，1982，55
[34] Lu Y N, Huang Zu－Qia, Peng H W. On the Laser Induced Synthesis of Dense Hydrogen. *Commun Theor Phys*, 1991, **15**：103
[35] Peng H W. A General Theory of Self－Consistent Field Beyond Hartree－Fock. *Commun Theor Phys*, 1993, **20**：239
[36] Peng H W, Chen Jian－Hua, Kuang Hui－Sun, *et al*. The Higher-Order Approximation in General Theory of Self－Consistent Field Beyond Hartree－Fock. *Commun Theor Phys*, 1995, **24**：139

其他论文

回忆

[37] 彭桓武. 水木常青，自强不息. 清华校友通讯，1981，4(3)：33
[38] 彭桓武. 易地三处的中学生涯. 科学家，1987年5月：33，44
[39] 彭桓武. 钱三强和我. 中国科学院院刊，1992，**7**(4)：34
[40] 彭桓武. 送别钱三强. 清华校友通讯，1992，**10**(26)：163
[41] 彭桓武. 忆玻恩、海特勒、薛定谔与我的几段谈话. 现代物理知识，1993，(6)：2
[42] 彭桓武. 谈谈我对钱三强先生组织领导我国科技工作的几点认识. 原子能科学技术，

1994,**28**(3):189
[43] 彭桓武.缅怀周培源老师(第三部分:治学精神).物理,1994,**23**(3):189
[44] 彭桓武.学生与先生.中国科学报,1994年11月30日
[45] 彭桓武.值年自白.清华校友通讯,1995,**4**(31):65
[46] 彭桓武.话说薛定谔、玻恩、海特勒.现代物理知识,1995 (1):35

物理学
[47] 彭桓武.谈物理学的性质和发展中的特点.物理,1981,**10**(3):147
[48] 彭桓武.创造性与物理思维.百科知识,1987(4):71
[49] 彭桓武.90年代物理学突破方向的展望.物理,1991,**20**(1):9
[50] 彭桓武.当代物理学的前沿发展与普物教学.物理通报,1991(11):6
[51] 彭桓武.对21世纪物理学的发展的一点猜想.物理,1993,**22**(3):129

交叉学科
[52] 彭桓武.化学物理学的诞生、发展及其研究意义.百科知识,1984(4):43
[53] 彭桓武.对理论生物物理探讨的初步认识.见:中国科学院《复杂性研究》编委会.复杂性研究.北京:科学出版社,1993.159